

77588

Schreiber, David

From: Zhou, Shubo (AU1631)
Sent: Tuesday, October 08, 2002 4:16 PM
To: Schreiber, David
Subject: seq search request

Hi David,

This is a Markush search and each SEQ ID is a very short peptide seq. Enjoy!

Joe

Shubo "Joe" Zhou, Ph.D.
Patent Examiner
(703)-605-1158
CM1/12D06
AU 1631, US PTO

Search Request

Requester's full name: Shubo "Joe" Zhou **Examiner #:** 78282

Art Unit: 1631 **Phone #:** 703-605-1158 **Mailbox #:** 12D01/CM1

Results format: pape

New Room #: 12D06

Serial #: 09/422,838

Please search:

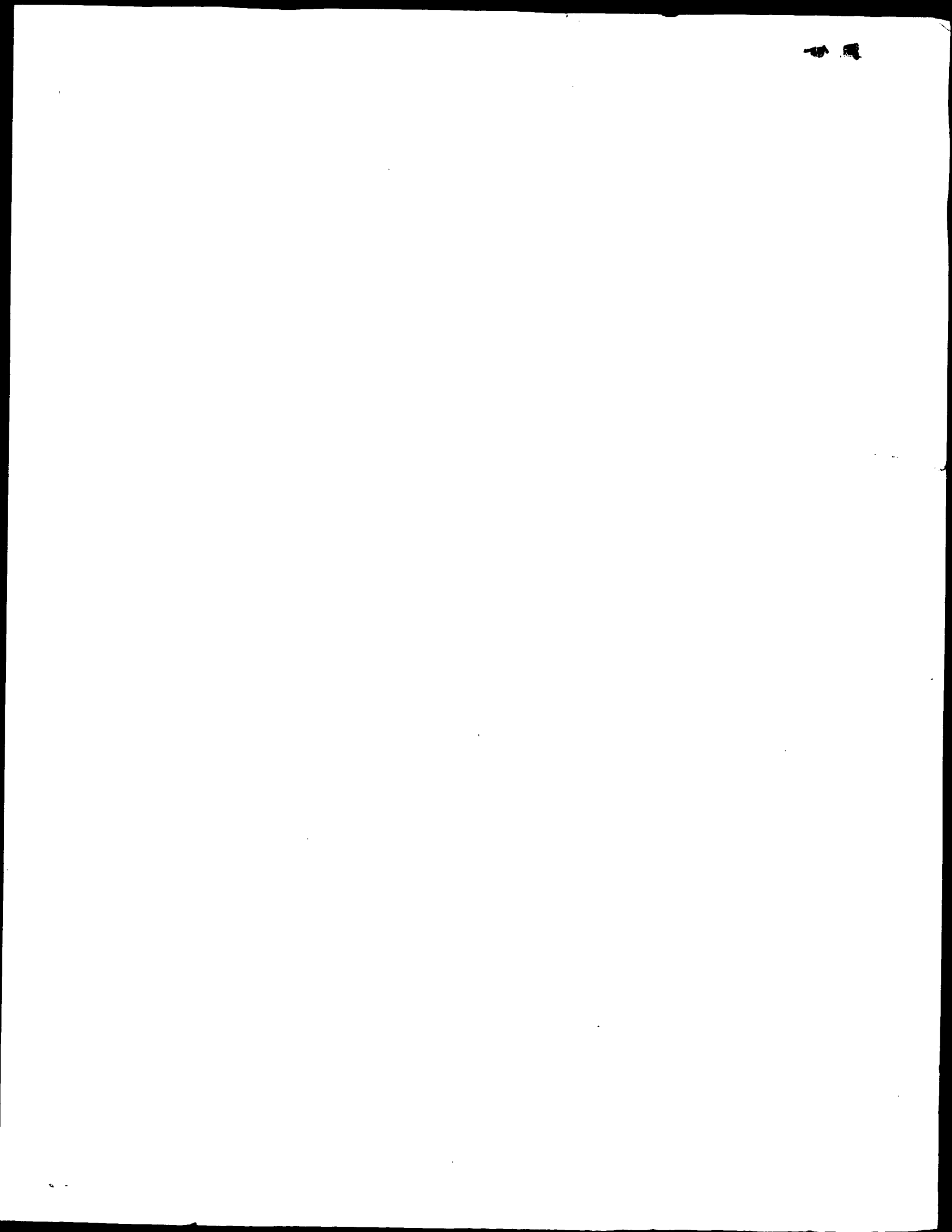
Protein databases for

SEQ ID NOs: 22-32, and 34

Including:

1. default search

Please provide 30 alignments for the search.



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 ; Search time 5.32084 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-22

Perfect score: 171

Sequence: 1 IEPTLRQWLAAARAGPNIEGPTLRQWLAAARA 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78.5	45.9	25	2	US-08-764-640-231
2	78.5	45.9	25	3	US-09-244-298A-231
3	78.5	45.9	25	4	US-09-516-704-231
4	73	42.7	14	2	US-08-764-640-13
5	73	42.7	14	2	US-08-764-640-193
6	73	42.7	14	3	US-08-973-225-13
7	73	42.7	14	3	US-08-973-225-193
8	73	42.7	14	3	US-09-244-298A-13
9	73	42.7	14	3	US-09-244-298A-193
10	73	42.7	14	4	US-09-516-704-13
11	73	42.7	14	4	US-09-516-704-193
12	73	42.7	15	2	US-08-764-640-17
13	73	42.7	15	2	US-08-764-640-185
14	73	42.7	15	3	US-08-973-225-17
15	73	42.7	15	3	US-08-973-225-185
16	73	42.7	15	3	US-09-244-298A-17
17	73	42.7	15	3	US-09-244-298A-185
18	73	42.7	15	4	US-09-516-704-17
19	73	42.7	15	4	US-09-516-704-185
20	73	42.7	16	2	US-08-764-640-18
21	73	42.7	16	2	US-08-764-640-194
22	73	42.7	16	2	US-08-764-640-232
23	73	42.7	16	3	US-08-973-225-18
24	73	42.7	16	3	US-08-973-225-194
25	73	42.7	16	3	US-08-973-225-220
26	73	42.7	16	3	US-09-244-298A-18
27	73	42.7	16	3	US-09-244-298A-194

28	73	42.7	16	3	US-09-244-298A-232	Sequence 232, App
29	73	42.7	16	4	US-09-516-704-18	Sequence 18, Appl
30	73	42.7	16	4	US-09-516-704-194	Sequence 194, App
31	73	42.7	16	4	US-09-516-704-232	Sequence 232, App
32	69	40.4	14	2	US-08-764-640-195	Sequence 195, App
33	69	40.4	14	2	US-08-764-640-199	Sequence 199, App
34	69	40.4	14	3	US-08-973-225-195	Sequence 195, App
35	69	40.4	14	3	US-08-973-225-199	Sequence 199, App
36	69	40.4	14	3	US-09-244-298A-195	Sequence 195, App
37	69	40.4	14	3	US-09-244-298A-199	Sequence 199, App
38	69	40.4	14	4	US-09-516-704-195	Sequence 195, App
39	69	40.4	14	4	US-09-516-704-199	Sequence 199, App
40	69	40.4	15	2	US-08-764-640-196	Sequence 196, App
41	69	40.4	15	2	US-08-764-640-200	Sequence 200, App
42	69	40.4	15	2	US-08-764-640-209	Sequence 209, App
43	69	40.4	15	2	US-08-764-640-215	Sequence 215, App
44	69	40.4	15	3	US-08-973-225-196	Sequence 196, App
45	69	40.4	15	3	US-08-973-225-200	Sequence 200, App

ALIGNMENTS

RESULT 1

US-08-764-640-231
; Sequence 231, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:

NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
US-08-764-640-231

Query Match 45.9%; Score 78.5; DB 2; Length 25;
Best Local Similarity 46.4%; Pred. No. 0.00027;
Matches 13; Conservative 8; Mismatches 2; Indels 5; Gaps 1;

QY 2 EGPTRLQWLAARAGPNEGPTLRQWLA 29
:||||:|:| :||||:|:|
Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 2

US-09-244-298A-231

; Sequence 231, Application US/09244298A
; Patent No. 6121238

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/244,298A

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 231:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 13

OTHER INFORMATION: /product= "Ava"

US-09-244-298A-231

Query Match 45.9%; Score 78.5; DB 3; Length 25;
Best Local Similarity 46.4%; Pred. No. 0.00027;
Matches 13; Conservative 8; Mismatches 2; Indels 5; Gaps 1;

QY 2 EGPTRLQWLAARAGPNEGPTLRQWLA 29

Db 2 DGPTLREWISFXA-----DGPTLREWIS 24
:||||:|:| :||||:|:|

RESULT 3

US-09-516-704-231

; Sequence 231, Application US/09516704
; Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 231:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 13

OTHER INFORMATION: /product= "Ava"

SEQUENCE DESCRIPTION: SEQ ID NO: 231:

US-09-516-704-231

Query Match 45.9%; Score 78.5; DB 4; Length 25;
Best Local Similarity 46.4%; Pred. No. 0.00027;

Matches 13; Conservative 8; Mismatches 2; Indels 5; Gaps 1;

QY 2 EGPTRLQWLAARAGPNEGPTLRQWLA 29

Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 4

US-08-764-640-13

; Sequence 13, Application US/08764640
; Patent No. 5869451

; Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 42.7%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
|||||
Db 1 IEPTLRQWLAARA 14

RESULT 5
US-08-764-640-193
Sequence 193, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193

Query Match 42.7%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
|||||
Db 1 IEPTLRQWLAARA 14

RESULT 6
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haseelden, Sherril S.
APPLICANT: Matheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A

FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 42.7%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 7
US-08-973-225-193
; Sequence 193, Application US/0897225A
; Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-08-973-225-193

Query Match 42.7%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 8
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 42.7%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 9
US-09-244-298A-193
; Sequence 193, Application US/09244298A

Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-193

Query Match 42.7%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEQPTLROWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEQPTLROWLAARA 14

RESULT 10
US-09-516-704-13
Sequence 13, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13

Query Match 42.7%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEQPTLROWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEQPTLROWLAARA 14

RESULT 11
US-09-516-704-193
Sequence 193, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match 42.7%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14

RESULT 12
US-08-764-640-17
; Sequence 17, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-185

Query Match 42.7%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 2 IEGPTLRQWLAAARA 15
```

```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-764-640-17

Query Match 42.7%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14

RESULT 13
US-08-764-640-185
; Sequence 185, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-185

Query Match 42.7%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 2 IEGPTLRQWLAAARA 15
```

RESULT 14
US-08-973-225-17
; Sequence 17, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-973-225-17

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
DB 1 IEPTLRQWLAAARA 14

RESULT 15
US-08-973-225-185
; Sequence 185, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.

Wagstrom, Christopher R.
Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEPTLRQWLAAARA 14
| | | | | | | | | | | | | | | |
DB 2 IEPTLRQWLAAARA 15

RESULT 16
US-09-244-298A-17
; Sequence 17, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-17

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARA 14
   |||||
DB 1 IEPTLRQWLAAARA 14

RESULT 17
US-09-244-298A-185
; Sequence 185, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-185

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARA 14
   |||||
DB 2 IEPTLRQWLAAARA 15

RESULT 18
US-09-516-704-17
; Sequence 17, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
; US-09-516-704-17

Query Match 42.7%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARA 14
   |||||
```

Db 1 IEGPTLRQWLAARA 14

RESULT 19

US-09-516-704-185

; Sequence 185, Application US/09516704

; Patent No. 5869451

; GENERAL INFORMATION:

APPLICANT: Dower, William J.

; Barrett, Ronald W.

; Cwirla, Steven E.

; Gates, Christian

; Schatz, Peter J.

; Balasubramanian, Palaniappan

; Wagstrom, Christopher R.

; Hendren, Richard W.

; Depvince, Randolph B.

; Podduturi, Surekha

; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-09-516-704-185

Query Match

Best Local Similarity 42.7%; Score 73; DB 4; Length 15;

Mismatches 0; Mismatches 0; Indels 0; Gaps 0;

Matches 14; Conservative 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14

Db 2 IEGPTLRQWLAARA 15

RESULT 20

US-08-764-640-18

; Sequence 18, Application US/08764640

; Patent No. 5869451

; Patent No. 5869451 5837683

; GENERAL INFORMATION:

APPLICANT: Dower, William J.

; Barrett, Ronald W.

; Cwirla, Steven E.

; Gates, Christian

; Schatz, Peter J.

; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

APPLICANT: Balasubramanian, Palaniappan

APPLICANT: Wagstrom, Christopher R.

APPLICANT: Hendren, Richard W.

APPLICANT: Depvince, Randolph B.

APPLICANT: Podduturi, Surekha

APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product= "Beta-ala"

US-08-764-640-18

Query Match

Best Local Similarity 42.7%; Score 73; DB 2; Length 16;

Mismatches 0; Mismatches 0; Indels 0; Gaps 0;

Matches 14; Conservative 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14

Db 1 IEGPTLRQWLAARA 14

RESULT 21

US-08-764-640-194

; Sequence 194, Application US/08764640

; Patent No. 5869451

; Patent No. 5869451 5837683

; GENERAL INFORMATION:

APPLICANT: Dower, William J.

; Barrett, Ronald W.

; Cwirla, Steven E.

; Gates, Christian

; Schatz, Peter J.

; Balasubramanian, Palaniappan

; Wagstrom, Christopher R.

; Hendren, Richard W.

; Depvince, Randolph B.

; Podduturi, Surekha

; Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-194

Query Match 42.7%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

RESULT 22
US-08-764-640-232
; Sequence 232, Application US/08764640
; Patent No. 5869451
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-232

Query Match 42.7%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

RESULT 23
US-08-973-225-18
; Sequence 18, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherrill S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
```


FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product="Beta-ala"
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 24
US-08-973-225-194
; Sequence 194, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

RESULT 25
US-08-973-225-220
; Sequence 220, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 220:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

RESULT 26
US-09-244-298A-18
; Sequence 18, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Depreince, Randolph B.

APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product- "Beta-ala"
US-09-244-298A-18

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
|||||
Db 1 IEGPTLRQWLAARA 14

RESULT 27
US-09-244-298A-194
Sequence 194, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
|||||
Db 2 IEGPTLRQWLAARA 15

RESULT 28
US-09-244-298A-232
Sequence 232, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-232

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
DB 2 IEGPTLRQWLAARA 15

RESULT 29

US-09-516-704-18
Sequence 18, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product= "Beta-ala"

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-516-704-18

Query Match 42.7%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
DB 1 IEGPTLRQWLAARA 14

RESULT 30

US-09-516-704-194

Sequence 194, Application US/09516704

Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwirla, Steven E.

Gates, Christian

Schatz, Peter J.

Balasubramanian, Palaniappan

Wagstrom, Christopher R.

Hendren, Richard W.

Deprince, Randolph B.

Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 194:

US-09-516-704-194

Query Match 42.7%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
DB 2 IEGPTLRQWLAARA 15

Search completed: October 9, 2002, 09:06:29

Job time : 6.32084 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 14.3888 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-22
Perfect score: 171
Sequence: 1 IEQPTLRQLWAARACPNIEGPTLRQLWAARA 32

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	100.0	32	21 AAB17297	TPO-mimetic peptid
2	171	100.0	32	21 AAY96520	Thrombopoietin mim
3	171	100.0	34	21 AAY96527	Thrombopoietin mim
4	156	91.2	32	21 AAB17289	TPO-mimetic peptid
5	147.5	86.3	31	21 AAB17288	TPO-mimetic peptid
6	147	86.0	30	21 AAB17287	TPO-mimetic peptid
7	145.5	85.1	33	21 AAB17290	TPO-mimetic peptid
8	145	84.8	34	21 AAB17291	TPO-mimetic peptid
9	145	84.8	36	21 AAB17306	TPO-mimetic peptid
10	145	84.8	36	21 AAY96526	Thrombopoietin mim
11	144.5	84.5	35	21 AAB17292	TPO-mimetic peptid

12 144 84.2 36 21 AAB16963 TPO-mimetic peptid
13 144 84.2 36 21 AAB17293 TPO-mimetic peptid
14 144 84.2 36 21 AAB17301 TPO-mimetic peptid
15 144 84.2 36 21 AAB17303 TPO-mimetic peptid
16 144 84.2 36 21 AAB17307 TPO-mimetic peptid
17 144 84.2 36 21 AAY96523 Thrombopoietin mim
18 144 84.2 36 21 AAY96524 Thrombopoietin mim
19 144 84.2 36 21 AAY96525 Thrombopoietin mim
20 144 84.2 41 21 AAY96528 Thrombopoietin mim
21 144 84.2 42 21 AAB17281 TPO-mimetic peptid
22 144 84.2 42 21 AAB17282 TPO-mimetic peptid
23 144 84.2 42 21 AAB17308 Thrombopoietin mim
24 144 84.2 42 21 AAY96530 Synthetic TMP-TMP
25 144 84.2 60 21 AAB17311 Synthetic TMP-TMP
26 144 84.2 269 21 AAB16960 TMP-TMP-Fc protein
27 144 84.2 269 21 AAY96531 Human IgG1 Fc TMP
28 143.5 83.9 37 21 AAB17294 TPO-mimetic peptid
29 143 83.6 38 21 AAB17295 TPO-mimetic peptid
30 142.5 83.3 39 21 AAB17304 TPO-mimetic peptid
31 142.5 83.3 39 21 AAB17305 TPO-mimetic peptid
32 142 83.0 40 21 AAB17302 TPO-mimetic peptid
33 141 82.5 42 21 AAB17296 TPO-mimetic peptid
34 140.5 82.2 29 21 AAB17286 Fc-TMP-TMP protein
35 140 81.9 268 21 AAB16959 TPO-mimetic peptid
36 134 78.4 28 21 AAB17285 TPO-mimetic peptid
37 133.5 78.1 29 21 AAB16970 TPO-mimetic peptid
38 133.5 78.1 31 21 AAB16973 TPO-mimetic peptid
39 133.5 78.1 31 21 AAB16974 TPO-mimetic peptid
40 127.5 74.6 29 21 AAB18971 TPO-mimetic peptid
41 120.5 70.5 29 21 AAB16975 TPO-mimetic peptid
42 120.5 70.5 29 21 AAB16976 TPO-mimetic peptid
43 118 69.0 36 21 AAB17298 TPO-mimetic peptid
44 118 69.0 36 21 AAB17299 TPO-mimetic peptid
45 118 69.0 36 21 AAY96521 Cyclic or linear t

ALIGNMENTS

RESULT 1

AAB17297
ID AAB17297 standard; Peptide; 32 AA.
XX
XX AAB17297;
DT 31-OCT-2000 (first entry)
XX TPO-mimetic peptide sequence SEQ ID NO:353.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; Epo; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0420802.

PA (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX PS Example 1; Page 316; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 31 AA;

Query Match 86.38; Score 147.5; DB 21; Length 31;

Best Local Similarity 93.8%; Pred. No. 3.8e-14;

Matches 30; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 IEGPTLRQWLAAARAGNGIEGPTLRQWLAAARA 32

|||||

Db 1 IEGPTLRQWLAAARAG-GGIEGPTLRQWLAAARA 31

|||||

RESULT 6

AAB17287

ID AAB17287 standard; Peptide; 30 AA.

XX AC AAB17287;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:343.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX KW vascular endothelial growth factor; matrix metalloproteinase;

XX KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX PS Example 1; Page 315-316; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX SQ Sequence 30 AA;

Query Match 86.0%; Score 147; DB 21; Length 30;

Best Local Similarity 93.8%; Pred. No. 4.3e-14;

Matches 30; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 IEGPTLRQWLAAARAGNGIEGPTLRQWLAAARA 32

|||||

Db 1 IEGPTLRQWLAAARAG-GGIEGPTLRQWLAAARA 30

|||||

RESULT 7

AAB17290

ID AAB17290 standard; Peptide; 33 AA.

XX AC AAB17290;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:346.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX KW vascular endothelial growth factor; matrix metalloproteinase;

XX KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 36 AA;

Query Match 84.8%; Score 145; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 1e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
 QY 1 IEPTLROWLAARA---GPNIGIEPTLROWLAARA 32
 DB 1 IEPTLROWLAARAAGGNGSGGIEPTLROWLAARA 36

RESULT 10

AA96526

ID AAY96526 standard; peptide; 36 AA.

XX AAY96526;

AC AAY96526;

XX 04-SEP-2000 (first entry)

DT Thrombopoietin mimetic peptide compound 7.

DE Thrombopoietin mimetic peptide compound 7.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;

XX anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;

XX immunosuppressive; anti-inflammatory; linker.

XX Synthetic.

XX Key

XX Modified-site 1

XX Location/Qualifiers

XX /note= "optionally linked to an Fc molecule"

XX 1..14

XX /label= TMP_1

XX 15..18

XX /label= linker

XX 19..32

XX /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the

XX production of platelets or platelet precursors, useful for treatment of

PT diseases which involve thrombocytopenia

XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin

XX mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2],

XX is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

XX 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,

XX X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and

XX X_1-X_1_4. X_1 = L, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;

XX X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,

XX or X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, L, V,

XX L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,

XX T, V, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker

XX comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

XX activate the c-Mpl receptor which mediates the activity of endogenous

XX thrombopoietin. The TMPs are useful for increasing the production of

XX platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which

XX is useful for treatment of diseases which involve thrombocytopenia, e.g.

XX aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency

XX virus associated ITP, and systemic lupus erythematosus.

XX

SQ Sequence 36 AA;

Query Match 84.8%; Score 145; DB 21; Length 36;

Best Local Similarity 83.3%; Pred. No. 1e-13;

Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLROWLAARA---GPNIGIEPTLROWLAARA 32

DB 1 IEPTLROWLAARAAGGNGSGGIEPTLROWLAARA 36

RESULT 11

AAB17292

ID AAB17292 standard; Peptide; 35 AA.

XX AAB17292;

AC AAB17292;

XX 31-OCT-2000 (first entry)

DT TPO-mimetic peptide sequence SEQ ID NO:348.

DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX vascular endothelial growth factor; matrix metalloproteinase;

XX asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

XX autoimmune diseases -

XX Example 1; Page 317-318; 608pp; English.

XX

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX

SQ Sequence 35 AA;
 Query Match 84.5%; Score 144.5; DB 21; Length 35;
 Best Local Similarity 85.7%; Pred. No. 1.2e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 1 IEGPTLRQWLAARA---GPNIGEGPTLRQWLAARA 32
 |||||
 DB 1 IEGPTLRQWLAARAAGGGGGGIEGPTLRQWLAARA 35
 |||||

RESULT 12
 AAB16963
 ID AAB16963 standard; Protein: 36 AA.
 XX
 AC AAB16963;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide TMP-TMP SEQ ID NO:14.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheatham J, Boone TC;
 XX
 DR WPI; 2000-350702/30.
 XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Disclosure; Page 190; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX

SQ Sequence 36 AA;
 Query Match 84.2%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAARA---GPNIGEGPTLRQWLAARA 32
 |||||
 DB 1 IEGPTLRQWLAARAAGGGGGGIEGPTLRQWLAARA 36
 |||||

RESULT 13
 AAB17293
 ID AAB17293 standard; Peptide: 36 AA.
 XX
 AC AAB17293;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:349.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheatham J, Boone TC;
 XX
 DR WPI; 2000-350702/30.
 XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 1; Page 318; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antitastmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Query Match 84.2%; Score 144; DB 21; Length 36;
Best Local Similarity 83.3%; Pred. No. 1.4e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1

QY 1 IEGPTLRQWLAARA----CPNGIEGPTLRQWLAARA 32

Db 1 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 15

AAB17303
 ID AAB17303 standard; peptide; 36 AA.

AC AAB17303:

XX

DT 31-OCT-2000 (first entry)

TP0-mimetic peptide sequence SEQ ID NO:359.

Modified peptide; therapeutic agent; fusion; FC domain; cancer.
 auto-immune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonists;
 MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 vascular endothelial growth factor; matrix metalloproteinase;
 asthma; thrombosis; pharmaceutical.

05 Synthetic

PN WO200024782-A2.

04-MAY-2000

25-OCT-1999: 99WO-US25044

23-OCT-1998: 98US-0105371-

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

Example 1: page 322: English:

The present invention describes composition of matter (1) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (1) is: (X1)a-P1-(X2)b, where: P1 = an Fc domain; X1 and X2 = are each independently selected from - (L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIGIEPTLRQWLAAARA 32
 |||||
 Db 1 IEGPTLRQWLAAARAAGGGGGGIEGPTLRQWLAAARA 36

RESULT 16

AAB17307
 ID AAB17307 standard; Peptide; 36 AA.

XX AAB17307;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:363.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 324; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIGIEPTLRQWLAAARA 32
 |||||
 Db 1 IEGPTLRQWLAAARAAGGGGGGIEGPTLRQWLAAARA 36

RESULT 17

AA96523

ID AA96523 standard; peptide; 36 AA.

XX AC AA96523;

DT 04-SEP-2000 (first entry)

DE Thrombopoietin mimetic peptide compound 4.

XX Thrombopoietin; mimetic; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14

FT Peptide 15..22 /label= TMP_1

FT Modified-site 18 /label= linker

FT Peptide 23..36 /note= "optionally modified by bromoacetyl or PEG"

FT /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,

CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4, X-1-I, A, V, L, S or R; X-2-E, D, K or V; X-3-G or A;
 CC X-4-P; X-5-T or S; X-6-L, I, V, A or F; X-7-R or K; X-8-Q, N,
 CC E; X-9-W, Y or F; X-10-L, I, V, A, F, M, or K; X-11-A, I, V,
 CC L, F, S, T, K, H, or E; X-12-A, I, V, L, F, G, S, or Q; X-13-R, K,
 CC T, V, N, Q or G; X-14-A, I, V, L, F, T, R, E, or G; L-1-linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 36 AA;
 CC Query Match 84.2%; Score 144; DB 21; Length 36;
 CC Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 CC Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA-----GPNIGIEGPTLRQWLAAARA 32
 DB 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 18
 AAY96524
 ID AAY96524 standard; peptide; 36 AA.

AC AAY96524;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 5.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX Synthetic.

XX Key Location/Qualifiers
 FT Modified-site 1 /note= "optionally linked to an Fc molecule"
 FT Peptide 1..14 /label= TWP_1
 FT Disulfide-bond 9..31 /note= "optional"
 FT Peptide 15..22 /label= linker
 FT Peptide 23..36 /label= TWP_2

XX WO200024770-A2.

PN 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 XX production of platelets or platelet precursors, useful for treatment of
 XX diseases which involve thrombocytopenia

PS Claim 16; Page 62; 91pp; English.

XX

CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1).TMP_2],
 CC is new TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4, X-1-I, A, V, L, S or R; X-2-E, D, K or V; X-3-G or A;
 CC X-4-P; X-5-T or S; X-6-L, I, V, A or F; X-7-R or K; X-8-Q, N,
 CC E; X-9-W, Y or F; X-10-L, I, V, A, F, M, or K; X-11-A, I, V,
 CC L, F, S, T, K, H, or E; X-12-A, I, V, L, F, G, S, or Q; X-13-R, K,
 CC T, V, N, Q or G; X-14-A, I, V, L, F, T, R, E, or G; L-1-linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 36 AA;

CC Query Match 84.2%; Score 144; DB 21; Length 36;
 CC Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 CC Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA-----GPNIGIEGPTLRQWLAAARA 32
 DB 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 19
 AAY96525

ID AAY96525 standard; peptide; 36 AA.

AC AAY96525;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 6.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.

XX Synthetic.

XX Key Location/Qualifiers
 FT Modified-site 1 /note= "optionally linked to an Fc molecule"
 FT Peptide 1..14 /label= TMP_1
 FT Peptide 15..18 /label= linker
 FT Peptide 19..32 /label= TMP_2
 FT Modified-site 32 /note= "optionally linked to an Fc molecule"

XX WO200024770-A2.

PN 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 XX production of platelets or platelet precursors, useful for treatment of

pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

XX PS Disclosure; Page 313; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

XX CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX CC where P1, P2, P3, and P4 = are each independently sequences of

XX CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX CC independently linkers; and a, b, c, d, e, and f = are each independently

XX CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX CC activities. DNAs, vectors and host cells from the present invention can

XX CC be used for producing pharmaceutical compositions. The compositions are

XX CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

XX CC The use of an Fc domain (rather than a Fab domain) can provide a longer

XX CC half-life or incorporate functions such as Fc receptor binding, protein

XX CC A binding, complement fixation, and possibly placental transfer. AAA69443

XX CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

XX CC sequences used in the exemplification of the present invention.

XX SQ Sequence 42 AA;

Query Match 84.2%; Score 144; DB 21; Length 42;

Best Local Similarity 83.3%; Pred. No. 1.7e-13;

Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAAARA----GPNIGIEGPTLRQWLAAARA 32

Db 7 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

RESULT 22

AAB17282

ID AAB17282 standard; Peptide; 42 AA.

XX AC AAB17282;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:338.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX KW vascular endothelial growth factor; matrix metalloproteinase;

XX KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and

XX PT pharmacologically active peptides, useful for treating cancer and

XX PT autoimmune diseases -

XX PS Disclosure; Page 313; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an

XX CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

XX CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

XX CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

XX CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

XX CC where P1, P2, P3, and P4 = are each independently sequences of

XX CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

XX CC independently linkers; and a, b, c, d, e, and f = are each independently

XX CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

XX CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

XX CC activities. DNAs, vectors and host cells from the present invention can

XX CC be used for producing pharmaceutical compositions. The compositions are

XX CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

XX CC The use of an Fc domain (rather than a Fab domain) can provide a longer

XX CC half-life or incorporate functions such as Fc receptor binding, protein

XX CC A binding, complement fixation, and possibly placental transfer. AAA69443

XX CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

XX CC sequences used in the exemplification of the present invention.

XX SQ Sequence 42 AA;

Query Match 84.2%; Score 144; DB 21; Length 42;

Best Local Similarity 83.3%; Pred. No. 1.7e-13;

Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAAARA----GPNIGIEGPTLRQWLAAARA 32

Db 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 23

AAB17308

ID AAB17308 standard; Peptide; 42 AA.

XX AC AAB17308;

XX DT 31-OCT-2000 (first entry)

XX DE Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

XX KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

XX KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

XX KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

XX KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

XX KW vascular endothelial growth factor; matrix metalloproteinase;

XX KW asthma; thrombosis; pharmaceutical.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and

XX PT pharmacologically active peptides, useful for treating cancer and

XX PT autoimmune diseases -

XX PS Example 2; Page 327; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antitasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 42 AA;

Query Match 84.2%; Score 144; DB 21; Length 42;
 Best Local Similarity 83.3%; Pred. No. 1.7e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIEGPTLRQWLAAARA 32
 DB 7 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42
 |||||

RESULT 24
 AAY96530

ID AAY96530 standard; Protein; 42 AA.

AC AAY96530;

DT 04-SEP-2000 (first entry)

DE Thrombopoietin mimetic peptide.

KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.

OS Synthetic.

PN WO200024770-A2.

PD 04-MAY-2000.

PF 22-OCT-1999; 99WO-US24834.

PR 23-OCT-1998; 98US-0105348.

PA (AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;

DR WPI; 2000-365108/31.

DR N-PSDB; AAA29225.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

PS Example 2a; Page 48; 91pp; English.

XX Overlapping oligonucleotides were used to construct a synthetic

CC gene encoding a thrombopoietin mimetic peptide (TMP), which
 CC was then fused in-frame to the Fc region of the human IgG1 chain (see
 CC AAY96529). A compound which binds to an mpl receptor comprising a TMP
 CC dimer joined by a linker [TMP-1-(L1)-nTMP-2], is new. TMP-1 and TMP-2
 CC are amino acid sequences varying from at least 10 to 14 residues in

CC length comprising X2-X1-0, X2-X1-1, X2-X1-2, X2-X1-3, X2-X1-4,
 CC X1-X1-0, X1-X1-1, X1-X1-2, X1-X1-3, and X1-X1-4. X1 = I, A,
 CC V, L, S or R; X2 = E, D, K or V; X3 = G or A; X4 = F; X5 = T or S;
 CC X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N, or E; X9 = W, Y or F;
 CC X10 = L, I, V, A, F, M, or K; X11 = A, I, V, L, F, S, T, K, H, or E;
 CC X12 = A, I, V, L, F, G, S, or Q; X13 = R, K, T, V, N, Q or G; X14 =
 CC A, I, V, L, F, T, E, or G; L1 = linker comprising 1 to 20 amino
 CC acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl
 CC receptor which mediates the activity of endogenous thrombopoietin. The
 CC TMPs are useful for increasing the production of platelets or platelet
 CC precursors (e.g. megakaryocytes) in a mammal, which is useful for
 CC treatment of diseases which involve thrombocytopenia, e.g. aplastic
 CC anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus
 CC associated ITP, and systemic lupus erythematosus.

XX Sequence 42 AA;

Query Match 84.2%; Score 144; DB 21; Length 42;
 Best Local Similarity 83.3%; Pred. No. 1.7e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIEGPTLRQWLAAARA 32
 DB 7 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42
 |||||

RESULT 25

AAB17311

ID AAB17311 standard; Peptide; 60 AA.

AC AAB17311;

DT 31-OCT-2000 (first entry)

DE Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

OS Homo sapiens.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

PS Example 2; Page 331; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 60 AA;

Query Match 84.2%; Score 144; DB 21; Length 60;
 Best Local Similarity 83.3%; Pred. NO. 2.5e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAARA----GPNIGEGPTLRQWLAARA 32
 |||||
 DB 2 IEPTLRQWLAARAAGGGGGGIEPTLRQWLAARA 37

RESULT 26

AAB16960
 ID AAB16960 standard; Protein; 269 AA.

XX AAB16960;

XX 31-OCT-2000 (first entry)

XX TNP-TMP-Fc protein sequence SEQ ID NO:10.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX N-PSDB; AAA69446.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases.

XX Example 2; Page 185-186; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 269 AA;

Query Match 84.2%; Score 144; DB 21; Length 269;
 Best Local Similarity 83.3%; Pred. NO. 1.4e-12;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAARA----GPNIGEGPTLRQWLAARA 32
 |||||
 DB 2 IEPTLRQWLAARAAGGGGGGIEPTLRQWLAARA 37

RESULT 27

AAY96531
 ID AAY96531 standard; Protein; 269 AA.

XX AAY96531;

XX 04-SEP-2000 (first entry)

XX Human IgG1 Fc TNP fusion protein.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TNP; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.

XX Homo sapiens.

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX N-PSDB; AAA29229.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Example 2A; Page 49-50; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP₁-(L₁)_n-TMP₂],
 CC is new. TMP₁ and TMP₂ are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X₂-X₁-L₀, X₂-X₁-L₁, X₂-X₁-L₂,
 CC X₂-X₁-L₃, X₂-X₁-L₄, X₁-X₁-L₀, X₁-X₁-L₁, X₁-X₁-L₂, X₁-X₁-L₃, and
 CC X₁-X₁-L₄. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K,
 CC T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 269 AA;

Query Match 84.2%; Score 144; DB 21; Length 269;
Best Local Similarity 83.3%; Pred. No. 1.4e-12;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAAAR-----GPNIGIEPTLRQWLAAAR 32
|||||
Db 234 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 269
|||||

RESULT 28

AAB17294

ID AAB17294 standard; Peptide; 37 AA.

XX AAB17294;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:350.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -

XX Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
XX activities. DNAs, vectors and host cells from the present invention can
XX be used for producing pharmaceutical compositions. The compositions can
XX be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
XX The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 37 AA;

Query Match 83.9%; Score 143.5; DB 21; Length 37;
Best Local Similarity 81.1%; Pred. No. 1.7e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

QY 1 IEPTLRQWLAAAR-----GPNIGIEPTLRQWLAAAR 32
|||||
Db 1 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 37
|||||

RESULT 29

AAB17295

ID AAB17295 standard; Peptide; 38 AA.

XX AAB17295;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:351.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
XX activities. DNAs, vectors and host cells from the present invention can
XX be used for producing pharmaceutical compositions. The compositions are
XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
XX The use of an Fc domain (rather than a Fab domain) can provide a longer
XX half-life or incorporate functions such as Fc receptor binding, protein
XX A binding, complement fixation, and possibly placental transfer. AAA69443

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-17 <NAD>
 A:Cross-references: GB:M33715
 R:Hester, K.; Luo, J.; Burns, G.; Braswell, E.H.; Sokatch, J.R.
 Eur. J. Biochem. 233, 828-836, 1995
 A:Title: Purification of active E1-alpha(2)-beta(2) of Pseudomonas putida branched-chain
 A:Reference number: S63475; MUID:96085147
 A:Accession: S63475
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-13 <HES>
 C:Genetics:
 A:Gene: bkdA1
 C:Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-bin
 C:Keywords: lipoamide; oxidoreductase; phosphoprotein; thiamin pyrophosphate
 F:2-410/Product: 3-methyl-2-oxobutanoate dehydrogenase (lipoamide) alpha chain #status F
 F:202-251/Domain: thiamin pyrophosphate-binding domain homology <TPB>
 F:313/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 29.2%; Score 50; DB 1; Length 410;
 Best Local Similarity 53.3%; Pred. No. 55;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 GPTLQWLAAARAGPN 17
 II: I : I : I : I : I :
 Db 298 GPSLIEWVTYRAGPH 312

RESULT 22
 C83365
 2-oxoisovalerate dehydrogenase (alpha subunit) PA2247 [imported] - Pseudomonas aeruginos
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: C83365
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: AB2950; MUID:20437337
 A:Accession: C83365
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-410 <STO>
 A:Cross-references: GB:AE004650; GB:AE004091; NID:g9948267; PIDN:AAG05635.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: bkdA1; PA2247
 C:Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-bin

Query Match 29.2%; Score 50; DB 2; Length 410;
 Best Local Similarity 53.3%; Pred. No. 55;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 GPTLQWLAAARAGPN 17
 II: I : I : I : I : I :
 Db 298 GPSLIEWVTYRAGPH 312

RESULT 23
 T38324
 probable trna methyltransferase - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T38324
 R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: Z21733
 A:Accession: T38324
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-415 <BRO>

A:Cross-references: EMBL:Z98977; PIDN:CAB11659.1; GSPDB:GN000066; SPDB:SPAC23H4.04
 A:Experimental source: strain 972h; cosmid c23H4
 C:Genetics:
 A:Gene: SPDB:SPAC23H4.04
 A:Map position: 1
 A:Introns: 34/1; 54/3
 C:Superfamily: probable membrane protein YDL033c

Query Match 29.2%; Score 50; DB 2; Length 415;
 Best Local Similarity 37.0%; Pred. No. 56;
 Matches 10; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 IEPTLQWLAAARAGPNIGPTLRQW 27
 II: I : I : I : I : I :
 Db 58 VEGVEMRWLDEDSAPSGC--PAERDW 82

RESULT 24
 S06469
 photosystem II chlorophyll a-binding protein psbc - Synechocystis sp. (strain PCC 680
 N:Alternate names: chlorophyll-binding protein, 43K; photosynthetic reaction center 4
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 07-Jun-1990 #sequence_revision 19-Jan-1996 #text_change 20-Jun-2000
 C:Accession: S06469; S07497; S02380; S74838
 R:Chisholm, D.; Williams, J.G.K.
 Plant Mol. Biol. 10, 293-301, 1988
 A:Title: Nucleotide sequence of psbc, the gene encoding the CP-43 chlorophyll a-bindi
 A:Reference number: S06469
 A:Accession: S06469
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 'MKTLSLRFRSPV', 2-460 <CHI>
 A:Cross-references: GB:M21538; NID:g340699; PIDN:AAA85378.1; PID:g1161272
 A:Note: this sequence uses an incorrect initiation codon
 R:Carpenter, S.D.; Charite, J.; Eggers, B.; Vermaas, W.F.J.
 FEBS Lett. 260, 135-137, 1990
 A:Title: The psbc start codon in Synechocystis sp. PCC 6803.
 A:Reference number: S07496; MUID:90127396
 A:Accession: S07497
 A:Molecule type: DNA
 A:Residues: 1-7 <CAR>
 A:Note: the authors definitively establish that the Met-1 GTG is the initiation codon a
 R:Dzelzkalns, V.A.; Bogorad, L.
 EMBO J. 7, 333-338, 1988
 A:Title: Molecular analysis of a mutant defective in photosynthetic oxygen evolution
 A:Reference number: S02379; MUID:88211542
 A:Accession: S02380
 A:Molecule type: DNA
 A:Residues: 'MKTLSLRFRSPV', 2-54, 'N', 56-149, 'I', 151-288 <DZE>
 A:Cross-references: EMBL:X07018; NID:g48064; PIDN:CAA30071.1; PID:g48066
 A:Note: the authors translated the codon CAT for residue 131 as Phe; this sequence us
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
 s.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S74838
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 'MKTLSLRFRSPV', 2-41, 'A', 43-460 <KAN>
 A:Cross-references: EMBL:D90909; GB:AB001339; NID:g152844; PIDN:BAAI799.1; PID:g165
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 A:Note: this sequence uses an incorrect initiation codon
 C:Genetics:
 A:Gene: psbc
 A:Start codon: GTG
 C:Superfamily: photosystem II chlorophyll a-binding protein psbc
 C:Keywords: chlorophyll; membrane-associated complex; photosynthesis; photosystem II;

Query Match 29.2%; Score 50; DB 2; Length 460;
 Best Local Similarity 35.0%; Pred. No. 62;

Matches 14; Conservative 3; Mismatches 13; Indels 10; Gaps 2;

QY 3 GPTLR-----OWLAARAGPNGIEGPTLRQ----WLAARA 32
| | | | |
Db 340 GETMRFWDFRGPWLPLRGPNGLDLKLRNDIQPWVRRRA 379

RESULT 25
T20454
hypothetical protein F01D4.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T20454
R:WILD, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19278
A:Accession: T20454
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-472 <WIL>
A:Cross-references: EMBL:Z81054; PIDN:CAB02881.1; GSPDB:GN00022; CESP:F01D4.4
A:Experimental source: clone F01D4
C:Genetics:
A:Gene: CESP:F01D4.4
A:Map position: 4
A:Introns: 59/3; 127/3; 334/3; 455/3
C:Superfamily: human carboxypeptidase H

Query Match 29.2%; Score 50; DB 2; Length 472;
Best Local Similarity 47.1%; Pred. No. 64;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 PTLROWLAARAGPNGIE 20
| | | | |
Db 157 PAQRQLWTGRSINGVD 173

RESULT 26
C70559
probable polA protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: C70559
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: C70559
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-904 <COL>
A:Cross-references: GB:Z95554; GB:AL123456; MID:g3261771; PIDN:CAB08882.1; PID:g2113913
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: polA
C:Superfamily: DNA-directed DNA polymerase I

Query Match 29.2%; Score 50; DB 2; Length 904;
Best Local Similarity 42.4%; Pred. No. 1.2e+02;
Matches 14; Conservative 1; Mismatches 8; Indels 10; Gaps 1;

QY 10 LAARAGPNGIEG-----PTLROWLAARA 32
| | | | |
Db 302 LAAAGGEVEDEGFDVRGCGALAPGTVRQWLAEHA 334

RESULT 27
A36925
transcription activator LysR-type Cbbr - Xanthobacter flavus
C:Species: Xanthobacter flavus

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
 A:Reference number: A97359; PMID:11743194
 A:Accession: H98202

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-355 <KUP>
 A:Cross-references: GB:AE007870; PIDN:AAK89146.1; PID:g15158956; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L1143
 A:Map position: linear chromosome

Query Match 28.9%; Score 49.5; DB 2; Length 355;
 Best Local Similarity 35.1%; Pred. No. 55;
 Matches 13; Conservative 4; Mismatches 13; Indels 7; Gaps 2;

QY 2 EGPTRLQWLAAR-----AGPNGIE--GPTLRQWLAAR 31
 Db 226 QQQSPFWIANMGEFVYPNGLERLAAQAARDWTAAR 262

RESULT 30

T22896

hypothetical protein F58B3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T22896

R:Harris, B.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z19633

A:Accession: T22896

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-214 <WIL>

A:Cross-references: EMBL:Z73427; PIDN:CAA97801.1; GSPDB:GN00022; CESP:F58B3.3

A:Experimental source: clone F58B3

C:Genetics:

A:Gene: CESP:F58B3.3

A:Map position: 4

A:Introns: 68/1

C:Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match 28.7%; Score 49; DB 2; Length 214;
 Best Local Similarity 50.0%; Pred. No. 38;
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 PTLRQWLAARAGPNGI 19

Db 191 PTHQWEGTAGPCGV 206

Search completed: October 9, 2002, 09:05:00
 Job time : 10.1944 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 3.82201 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-22

Perfect score: 171

Sequence: 1 IEPTLRQWLAARAGPNIGPTLRQWLAARA 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.5	31.3	266	1 SCO2_HUMAN	O43819 homo sapien
2	53	31.0	524	1 VGLG_CHAV	P13180 chandipura
3	52.5	30.7	814	1 AD15_HUMAN	Q13444 homo sapien
4	51	29.8	696	1 SP15_TORCA	P19965 torpedo cal
5	50	29.2	243	1 DEFH_HUMAN	Q9hbh1 homo sapien
6	50	29.2	246	1 TONB_PASHA	P72204 pasteurella
7	50	29.2	410	1 ODBA_PSEPU	P09060 pseudomonas
8	50	29.2	415	1 TRMU_SCHPO	O13947 schizosacch
9	50	29.2	472	1 PSBC_SYNY3	P09193 synchocyst
10	50	29.2	904	1 DPOI_MYCTU	Q07700 mycobacteri
11	49.5	28.9	333	1 CBBR_XANFL	P25545 xanthobacte
12	49	28.7	368	1 ODBA_BAGST	P21873 bacillus st
13	49	28.7	385	1 DIAC_HUMAN	Q01459 homo sapien
14	49	28.7	735	1 CNGI_CHICK	Q08005 gallus gall
15	49	28.7	911	1 CALB_BOVIN	Q28083 bos taurus
16	48.5	28.4	122	1 UROC_MOUSE	P81615 mus musculu
17	48	28.1	72	1 VAXIS_BP434	P11683 bacterioph
18	48	28.1	72	1 VAXIS_LAMBD	P03699 bacterioph
19	48	28.1	270	1 YL76_VIBCH	Q9kb28 vibrio chol
20	48	28.1	297	1 XERC_MYCLE	Q9cbu0 mycobacteri
21	48	28.1	370	1 ODBA_BACSU	P21881 bacillus su
22	48	28.1	1366	1 CA21_HUMAN	P08123 homo sapien
23	47.5	27.8	562	1 SYK_AERPE	Q9yft9 aeropyrum p
24	47	27.5	113	1 FRT2_HUMAN	O75474 homo sapien
25	47	27.5	357	1 PYRD_MYCTU	O06236 mycobacteri
26	47	27.5	473	1 PSBC_PINTH	P41643 pinus thunb
27	47	27.5	1338	1 PURA_HUMAN	O15067 homo sapien
28	47	27.5	1372	1 CA21_MOUSE	Q01149 mus musculu
29	47	27.5	1446	1 IE18_PPRVA	P33479 pseudorabie
30	47	27.5	1461	1 IE18_PPRVF	P11675 pseudorabie
31	47	27.5	1711	1 PTPQ_RAT	Q64612 rattus norv
32	47	27.5	1806	1 CALB_HUMAN	P12107 homo sapien
33	46	26.9	298	1 XERC_MYCTU	Q10815 mycobacteri

ALIGNMENTS

RESULT 1

ID	SCO2_HUMAN	STANDARD;	PRT;	266 AA.
AC	O43819; O9UK87;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	SCO2 protein homolog, mitochondrial precursor.			
GN	SCO2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Monocytes;			
RA	Smink L.J., Burton J.;			
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A., AND VARIANTS FTC LYS-140 AND PHE-225.			
RX	MEDLINE=20014747; PubMed=10545952;			
RA	Papadopoulos L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,			
RA	Sadlock J.E., Krishna S., Walker W., Selby J., Glerum D.M.,			
RA	Van Coster R., Lyon G., Scalsis E., Lebel R., Kaplan P., Shanske S.,			
RA	De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.;			
RT	*Fatal infantile cardioencephalomyopathy with COX deficiency and			
RT	mutations in SCO2, a COX assembly gene.*;			
RL	Nat. Genet. 23:333-337(1999).			
CC	-1- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER			
CC	TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.			
CC	-1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).			
CC	-1- TISSUE SPECIFICITY: UBIQUITOUS.			
CC	-1- DISEASE: DEFECTS IN SCO2 ARE THE CAUSE OF FATAL INFANTILE			
CC	CARDIOENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS			
CC	CHARACTERIZED BY HYPERPHOSPHATASE, LACTIC ACIDOSIS, AND			
CC	GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX			
CC	ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX			
CC	DEFICIENCIES.			
CC	-1- SIMILARITY: BELONGS TO THE SCO1/2 FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; AF177385; AAF05313.1; -			
DR	EMBL; AL021683; CAA16671.1; -			
DR	MIM; 604272; -			
DR	MIM; 604377; -			
DR	MIM; 220110; -			
DR	InterPro; IPR003782; SCO1_SenC.			
DR	Pfam; PF02630; SCO1_SenC; 1.			
KW	Mitochondrion; Transit peptide; Disease mutation; Polymorphism.			

O06399 mycobacteri
P02460 gallus gall
P49471 odontella s
P54234 clarkia arc
P54236 clarkia fra
P54239 clarkia wil
P54240 clarkia xan
P54243 oenothera m
P54235 clarkia con
P34796 clarkia lew
P54238 clarkia ros
P23608 a poly-beta

34 46 26.9 335 1 FABH_MYCTU
35 46 26.9 369 1 CA12_CHICK
36 46 26.9 509 1 PSBB_ODOSI
37 46 26.9 568 1 G6PI_CLAAR
38 46 26.9 568 1 G6PI_CLAFR
39 46 26.9 568 1 G6PI_CLAWI
40 46 26.9 568 1 G6PI_CLAXA
41 46 26.9 568 1 G6PI_OENME
42 46 26.9 569 1 G6PI_CLACO
43 46 26.9 569 1 G6PI_CLALE
44 46 26.9 570 1 G6PI_CLARO
45 46 26.9 589 1 PHBC_ALCEU

FT	TRANSIT	1	41	MITOCHONDRION (POTENTIAL).
FT	CHAIN	42	266	SCO2 PROTEIN HOMOLOG.
FT	VARIANT	20	20	R -> P (IN DBSNP:140523).
FT				/FTID=VAR_0111738.
FT	VARIANT	140	140	E -> K (IN FIC).
FT				/FTID=VAR_008874.
FT	VARIANT	225	225	S -> F (IN FIC).
FT				/FTID=VAR_008875.
FT	SEQUENCE	266 AA;	29810 MW;	BC2F40E057329BF3 CRC64;
Query Match 31.3%; Score 53.5; DB 1; Length 266;				
Best Local Similarity 33.3%; Pred. No. 4.8;				
Matches 16; Conservative 2; Mismatches 9; Indels 21; Gaps 2;				
QY	6	LOWLAARAGP-----NIEGPTLR-----	OWLAARA 32	
DB	33	LRSWLLSRQGPAGTGGQPGQGLRLLITGLFLGAGLGAWLAIRA 80		
RESULT 2				
VGUG_CHAV	STANDARD;	PRT;	524 AA.	
ID	VGUG_CHAV			
AC	PL180;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Spike glycoprotein precursor.			
GN	G.			
OS	Chandipura virus (strain I653514).			
OC	Viruses; ssRNA negative-strand viruses; Mononegavirales;			
OC	Rhabdoviridae; Vesiculovirus.			
OX	NCBI_TaxID=11273;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89299473; PubMed=2741347;			
RA	Masters P.S., Shella R.S., Butcher M., Patel B., Ghosh H.P.,			
RA	Banerjee A.K.;			
RT	"Structure and expression of the glycoprotein gene of Chandipura			
RT	virus.";			
RL	Virology 171:285-290(1989).			
CC	-1- FUNCTION: THIS PROTEIN FORMS SPIKES ON THE SURFACE OF THE VIRION.			
CC	IT IS RESPONSIBLE BOTH FOR THE BINDING OF THE VIRUS TO SUSCEPTIBLE			
CC	HOST CELLS AND FOR INDUCING THE UPTAKE OF THE VIRUS BY THE CELL.			
CC	THE INTERACTION BETWEEN THE INTERNAL COMPONENTS OF THE VIRION			
CC	AND THE PORTION OF THE GLYCOPROTEIN EXPOSED ON THE CYTOPLASMIC			
CC	FACE OF THE PLASMA MEMBRANE PROBABLY DIRECTS ENVELOPMENT AND			
CC	VIRUS BUDDING.			
CC	-1- SUBUNIT: TRIMERS IN THE ENDOPLASMIC RETICULUM.			
CC	-1- PTM: THIS PROTEIN IS MODIFIED BY THE COVALENT ADDITION OF PALMITIC			
CC	ACID VIA A THIOETHER LINKAGE TO A CYSTEINE. IT COULD BE EITHER ON			
CC	POSITION 479 OR 484.			
CC	-1- SIMILARITY: 39% IDENTITY TO THE G PROTEINS OF VSV.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; J04350; AAA2916.1; --			
DR	PIR; A32443; VGVNCV.			
DR	InterPro; IPR001903; Rhabd_glycop.			
DR	Pfam; PF00974; Rhabd_glycop; 2.			
KW	Transmembrane; Envelope protein; Glycoprotein; Lipoprotein; Palmitate;			
KW	Signal.			
FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	524	SPIKE GLYCOPROTEIN.
FT	DOMAIN	22	472	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	473	496	POTENTIAL.
FT	DOMAIN	497	524	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	184	184	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	344	344	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	LIPID	479	479	PALMITATE (POTENTIAL).
FT	LIPID	484	484	PALMITATE (POTENTIAL).
SO	SEQUENCE	524 AA;	58826 MW;	A84AF0A5FFFB73CD CRC64;
Query Match 31.0%; Score 53; DB 1; Length 524;				
Best Local Similarity 37.0%; Pred. No. 11;				
Matches 10; Conservative 4; Mismatches 13; Indels 0; Gaps 0;				
QY	1	IEGPTLRQWLAARAGPNGIEGPTLRQW 27		
DB	359	IDGVLKEPKRGKRESPSGISSDIWTQW 385		
RESULT 3				
AD15_HUMAN	STANDARD;	PRT;	814 AA.	
ID	AD15_HUMAN			
AC	Q13444; Q13493;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	ADAM 15 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase			
DE	domain 15) (Metalloproteinase-like, disintegrin-like, and cysteine-			
DE	rich protein 15) (MDC-15) (Metalloprotease RGD disintegrin protein)			
DE	(Metargidin).			
GN	ADAM15 OR MDC15.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE=Breast carcinoma;			
RC	MEDLINE=96214870; PubMed=8617717;			
RA	Kraetzschmar J., Lum L., Blobel C.P.;			
RT	"Metargidin, a membrane-anchored metalloprotease-disintegrin protein			
RT	with an RGD integrin binding sequence.";			
RL	J. Biol. Chem. 271:4593-4598(1996).			
CC	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Umbilical vein;			
RA	MEDLINE=97192141; PubMed=9039960;			
RX	Herren B., Raines E.W., Ross R.;			
RT	"Expression of a disintegrin-like protein in cultured human vascular			
RT	cells and in vivo.";			
RL	FASEB J. 11:173-180(1997).			
CC	[3]			
RP	INTERACTION WITH INTEGRIN ALPHA-V-BETA3.			
RX	MEDLINE=98184837; PubMed=9516430;			
RA	Zhang X.P., Kamata T., Yokoyama K., Puzon-McLaughlin W., Takada Y.;			
RT	"Specific interaction of the recombinant disintegrin-like domain of			
RL	MDC-15 (metargidin, ADAM-15) with integrin alphavbeta3.";			
CC	J. Biol. Chem. 273:7345-7350(1998).			
CC	-1- FUNCTION: MAY BE INVOLVED IN CELL-SURFACE PROTEOLYSIS, CELL			
CC	ADHESION OR INTRACELLULAR PROTEIN MATURATION.			
CC	-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	-1- SUBUNIT: INTERACTS WITH INTEGRIN ALPHA-V-BETA3, ENDOPHLIN I AND			
CC	SORTING NEXIN 9. ENDOPHLIN I AND SORTING NEXIN 9 PREFERENTIALLY			
CC	BIND THE PRECURSOR BUT NOT THE PROCESSED FORM OF ADAM15, THWAY			
CC	SUGGESTING THAT THE INTERACTION OCCURS IN A SECRETORY PATHWAY			
CC	COMPARTMENT PRIOR TO THE MEDIAL GOLGI (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED. OVEREXPRESSED IN			
CC	ARTHEROSCLEROTIC LESIONS. CONSTITUTIVELY EXPRESSED IN CULTURED			
CC	ENDOTHELIUM AND SMOOTH MUSCLE.			
CC	-1- DOMAIN: THE CYTOPLASMIC DOMAIN INTERACTS WITH ENDOPHLIN I AND			
CC	SORTING NEXIN 9 (BY SIMILARITY).			
CC	-1- DOMAIN: DESINTEGRIN DOMAIN BINDS TO INTEGRIN ALPHA-V-BETA3.			
CC	-1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY			
CC	SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.			
CC	-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.			

Q9HBH1; 01-MAR-2002 (Rel. 41, Created)
 01-MAR-2002 (Rel. 41, Last sequence update)
 01-MAR-2002 (Rel. 41, Last annotation update)
 Peptide deformylase, mitochondrial precursor (EC 3.5.1.88) (PDF)
 (Polypeptide deformylase).
 PDF1A OR PDF.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxId=9606;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=70514156; PubMed=11060042;
 Gigliione C., Serero A., Pierre M., Boisson B., Meinel T.;
 "Identification of eukaryotic peptide deformylases reveals
 universality of N-terminal protein processing mechanisms.";
 EMBO J. 19:5916-5929(2000).
 [2]
 SEQUENCE FROM N.A.
 Lonetto M.A., Zhu Y., Li X., Southan C.;
 "A human homolog of bacterial peptide deformylases.";
 Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: Removes the formyl group from the N-terminal Met of
 newly synthesized proteins (By similarity).
 -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
 methionyl peptide.
 -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
 -!- SUBCELLULAR LOCATION: Mitochondrion (Potential).
 -!- TISSUE SPECIFICITY: Ubiquitous.
 -!- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 EMBL; AF239156; AAC33968.1; -;
 EMBL; AF322879; AAK15624.1; -;
 InterPro; IPR000181; Pep_deformylase.
 Pfam; PF01327; Pep_deformylase; 1.
 ProDom; PD003844; Pep_deformylase; 1.
 Protein biosynthesis; Hydrolase; Iron; Mitochondrion; Transit peptide.
 TRANSIT 1 ?
 CHAIN ? 243 PEPTIDE DEFORMYLASE.
 METAL 172 172 IRON (BY SIMILARITY).
 METAL 214 214 IRON (BY SIMILARITY).
 ACT_SITE 215 215 BY SIMILARITY.
 METAL 218 218 IRON (BY SIMILARITY).
 SEQUENCE 243 AA; 27013 MW; B15A3456F0F8D689 CRC64;
 Query Match 29.2%; Score 50; DB 1; Length 243;
 Best Local Similarity 50.0%; Pred. No. 12;
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 QY 11 AARAGPNGIEGPTLRQ 26
 : : | : | : | : | : | :
 DB 31 SSTAAPDVGEGPALRR 46
 RESULT 6
 TONB_PASHA
 ID TONB_PASHA STANDARD; PRT; 246 AA.
 AC P72204;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE TonB protein.
 GN TONB.
 OS Pasteurella haemolytica.

```

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Mannheimia.
OC NCBI_TaxID=75985;
[1]
OC SEQUENCE FROM N.A.
OC STRAIN-SEROTYPE A1 / ATCC 43270;
OC Graham M.R., Lo R.Y.C.;
OC Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
OC -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
OC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
OC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
OC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
OC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
OC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
OC MEMBRANE PROTEINS (BY SIMILARITY).
OC -!- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
OC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
OC PERIPLASM (BY SIMILARITY).
OC -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
OC -----
OC This SWISS-PROT entry is copyright. It is produced through a collaboration
OC between the Swiss Institute of Bioinformatics and the EMBL outstation -
OC the European Bioinformatics Institute. There are no restrictions on its
OC use by non-profit institutions as long as its content is in no way
OC modified and this statement is not removed. Usage by and for commercial
OC entities requires a license agreement (See http://www.isb-sib.ch/announce/
OC or send an email to license@isb-sib.ch).
OC -----
OC EMBL; U62565; AAB09530.1; -.
OC Transport; Protein transport; Inner membrane; Periplasmic;
OC Transmembrane; Signal-anchor.
OC DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
OC FT FT 8 28 SIGNAL-ANCHOR (POTENTIAL).
OC FT FT 29 246 PERIPLASMIC (POTENTIAL).
OC SEQUENCE 246 AA; 27785 MW; C9582P619FCBA5B5 CRC64;
OC
OC Query Match 29.2%; Score 50; DB 1; Length 246;
OC Best Local Similarity 47.4%; Pred. No. 13;
OC Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
OC
OC QY 3 GPTLRQWLAAARAGPNIGEG 21
OC ||::||::||::||::||
OC DB 157 GPEIKQGIIVAKAIIPNAEAG 175
OC
OC RESULT 7
OC ID ODBA.PSEPU STANDARD; PRT; 410 AA.
OC AC P09060.
OC DT 01-NOV-1988 (Rel. 09, Created)
OC DT 01-FEB-1996 (Rel. 33, Last sequence update)
OC DT 30-MAY-2000 (Rel. 39, Last annotation update)
OC DE 2-oxoisovalerate dehydrogenase alpha subunit (EC 1.2.4.4) (Branched-
OC DE chain alpha-keto acid dehydrogenase component alpha chain (E1))
OC DE (BCKDH E1-alpha).
OC GN BKDAL.
OC OS Pseudomonas putida.
OC OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC OC Pseudomonas.
OC ON NCBI_TaxID=303;
OC RN [1]
OC RP SEQUENCE FROM N.A.
OC RC STRAIN=PPG2;
OC RX MEDLINE=88329084; PubMed=3416875;
OC RA Burns G., Brown T., Hatter K., Idriss J., Sokatch J.R.;
OC RT "Similarity of the E1 subunits of branched-chain-oxoacid dehydrogenase
OC RT from Pseudomonas putida to the corresponding subunits of mammalian
OC RL branched-chain-oxoacid and pyruvate dehydrogenases.";
OC RL Eur. J. Biochem. 176:311-317(1988).
OC [2]
OC RN SEQUENCE OF 1-17 FROM N.A.
OC RP STRAIN=PPG2;
OC RX MEDLINE=91008935; PubMed=2211503;

```


RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.:
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY.
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA(N)]
CC -I- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L11920; AAB46393.1; -
CC EMBL; Z95554; CAB08882.1; -
CC EMBL; AE007030; AAK45935.1; -
CC HSP; P19821; IBGX.
CC TIGR; MT1665; -
CC TubercuList; Rv1629; -
CC InterPro; IPR002562; 3_5_exonuclease.
CC InterPro; IPR002421; 5_3_exonuclease.
CC InterPro; IPR002298; DNA_pol.
CC InterPro; IPR001098; DNA_pol_A.
CC InterPro; IPR000513; Exo_N.1.
CC InterPro; IPR003583; HHH_1.
CC InterPro; IPR003584; HHH_2.
CC Pfam; PF01367; 5_3_exonuclease; 1.
CC Pfam; PF02739; 5_3_exonuc.N; 1.
CC Pfam; PF00476; DNA_pol_A; 1.
CC PRINTS; PR00868; DNAPOLI.
CC SMART; SM00474; 53EXOC; 1.
CC SMART; SM00475; 53EXOC; 1.
CC SMART; SM00278; Hhh1; 1.
CC SMART; SM00279; Hhh2; 1.
CC SMART; SM00482; POLAC; 1.
CC PROSITE; PS00447; DNA_POLYMERASE_A; 1.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
KW NCBI_TaxID=1773;
SQ
Query Match 29.2%; Score 50; DB 1; Length 904;
Best Local Similarity 35.0%; Pred. No. 24;
Matches 14; Conservative 3; Mismatches 13; Indels 10; Gaps 2;
QY 3 GPTLR-----QWLAAGPNGIEGPTLRQ-----WLAARA 32
DB 352 GETMREDFRGWLEPGRGPNGLDLKLRNDIQPWVRRRA 391
RESULT 10
DPOL_MYCTU STANDARD; PRT; 904 AA.
AC Q07700;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA OR Rv1629 OR MT1665 OR MTCV01B2.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=94124016; PubMed=8294019;
RA Mizrahi V., Huberts P., Dawes S.S., Dudding L.R.;
RT "A PCR method for the sequence analysis of the gyrA, polA and rnhA
RL gene segments from mycobacteria.";
RN Gene 136:287-290(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Hollroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA RA
RA Bishai W.:
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY.
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA(N)]
CC -I- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L11920; AAB46393.1; -
CC EMBL; Z95554; CAB08882.1; -
CC EMBL; AE007030; AAK45935.1; -
CC HSP; P19821; IBGX.
CC TIGR; MT1665; -
CC TubercuList; Rv1629; -
CC InterPro; IPR002562; 3_5_exonuclease.
CC InterPro; IPR002421; 5_3_exonuclease.
CC InterPro; IPR002298; DNA_pol.
CC InterPro; IPR001098; DNA_pol_A.
CC InterPro; IPR000513; Exo_N.1.
CC InterPro; IPR003583; HHH_1.
CC InterPro; IPR003584; HHH_2.
CC Pfam; PF01367; 5_3_exonuclease; 1.
CC Pfam; PF02739; 5_3_exonuc.N; 1.
CC Pfam; PF00476; DNA_pol_A; 1.
CC PRINTS; PR00868; DNAPOLI.
CC SMART; SM00474; 53EXOC; 1.
CC SMART; SM00475; 53EXOC; 1.
CC SMART; SM00278; Hhh1; 1.
CC SMART; SM00279; Hhh2; 1.
CC SMART; SM00482; POLAC; 1.
CC PROSITE; PS00447; DNA_POLYMERASE_A; 1.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
KW NCBI_TaxID=1773;
SQ
Query Match 29.2%; Score 50; DB 1; Length 472;
Best Local Similarity 35.0%; Pred. No. 24;
Matches 14; Conservative 3; Mismatches 13; Indels 10; Gaps 2;
QY 3 GPTLR-----QWLAAGPNGIEGPTLRQ-----WLAARA 32
DB 352 GETMREDFRGWLEPGRGPNGLDLKLRNDIQPWVRRRA 391
RESULT 10
DPOL_MYCTU STANDARD; PRT; 904 AA.
AC Q07700;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA OR Rv1629 OR MT1665 OR MTCV01B2.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=94124016; PubMed=8294019;
RA Mizrahi V., Huberts P., Dawes S.S., Dudding L.R.;
RT "A PCR method for the sequence analysis of the gyrA, polA and rnhA
RL gene segments from mycobacteria.";
RN Gene 136:287-290(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Hollroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,


```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M95767; AAC35684.1;
DR EMBL: AF085706; AAC35852.1;
DR EMBL: AF085700; AAC35852.1; JOINED.
DR EMBL: AF085701; AAC35852.1; JOINED.
DR EMBL: AF085702; AAC35852.1; JOINED.
DR EMBL: AF085703; AAC35852.1; JOINED.
DR EMBL: AF085704; AAC35852.1; JOINED.
DR EMBL: AF085705; AAC35852.1; JOINED.
DR PIR: A44102; A44102.
DR PIR: S27959; S27959.
DR MIM: 600873;
DR InterPro: IPR001579; Chitinase.2.
DR PROSITE: PS01095; CHITINASE.18; 1.
KW Hydrolase; Glycosidase; Signal; Lysosome; Glycoprotein.
FT SIGNAL 1 38 BY SIMILARITY.
FT CHAIN 39 385 DI-N-ACETYLCHITOBIASE.
FT ACT_SITE 143 143 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 193 193 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 385 AA; 43759 MW; 0A9D14C8B26B52EE CRC64;

Query Match 28.7%; Score 49; DB 1; Length 385;
Best Local Similarity 37.9%; Pred. No. 27;
Matches 11; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 4 PTLROWLAARAGPNIGEGPTLROWLAARA 32
| | | | : | | | : | | |
DB 4 PQLRRRLVSSPPSGVPGGLALLALLALLA 32

RESULT 14
CNGL_CHICK
ID CNGL_CHICK STANDARD; PRT; 735 AA.
AC Q90805;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cyclic nucleotide gated channel, cone photoreceptor, alpha subunit
DE (CNG channel 1) (CNG-1).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93264082; PubMed=7684234;
RA Boenigk W., Altenhofen W., Mueller F., Dose A., Illing M.,
RA Molday R.S., Kaupp U.B.;
RT "Rod and cone photoreceptor cells express distinct genes for
RT cGMP-gated channels.";
RL Neuron 10:865-877(1993).
CC -1- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF CONE
CC PHOTORECEPTORS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: X89598; CAA61757.1;
DR InterPro: IPR000636; Cation_chan_non_lig.
DR InterPro: IPR000595; cNMP_binding.
DR Pfam: PF00027; cNMP_binding; 1.
DR Pfam: PF00520; ion_trans; 1.
DR SMART: SM00100; cNMP; 1.
DR PROSITE: PS00888; cNMP_BINDING.1; 1.
DR PROSITE: PS00889; cNMP_BINDING.2; 1.
DR PROSITE: PS50042; cNMP_BINDING.3; 1.
KW Ionic channel; Ion transport; cAMP-binding; Transmembrane; Vision;
KW Multigene family.
FT DOMAIN 1 210 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 211 230 H1 (POTENTIAL).
FT DOMAIN 231 243 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 244 262 H2 (POTENTIAL).
FT DOMAIN 263 286 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 287 306 H3 (POTENTIAL).
FT DOMAIN 307 344 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 345 367 H4 (POTENTIAL).
FT DOMAIN 368 419 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 420 439 H5 (POTENTIAL).
FT DOMAIN 440 523 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 524 544 H6 (POTENTIAL).
FT DOMAIN 545 735 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 532 654 CAMP (BY SIMILARITY).
FT BINDING 591 591 CAMP (POTENTIAL).
FT BINDING 606 606 CAMP (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 735 AA; 85031 MW; A67ADFDD942CEFC CRC64;

Query Match 28.7%; Score 49; DB 1; Length 735;
Best Local Similarity 34.1%; Pred. No. 51;
Matches 14; Conservative 2; Mismatches 11; Indels 14; Gaps 2;

QY 1 IEGPTL-----RWLAARAGPNIGEGPTLROWLAAR 31
| | | | : | | | : | | |
DB 103 IRGPVELVSSRQSNIRSFQIGREQGVGVP-----WPLAR 139

RESULT 15
CALB_BOVIN
ID CALB_BOVIN STANDARD; PRT; 911 AA.
AC Q28083;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Collagen alpha 1(XI) chain (Fragment).
GN COL1A1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Smooth muscle;
RX MEDLINE=92078200; PubMed=1744123;
RA Brown K.E., Lawrence R., Sonenshein G.E.;
RA "Concerted modulation of alpha 1(XI) and alpha 2(V) collagen mRNAs in
RT bovine vascular smooth muscle cells.";
RL J. Biol. Chem. 266:23268-23273(1991).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD
CC OF ALPHA 3(XI)=1(XI) (BY SIMILARITY).
CC -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
CC -1- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```


RP SEQUENCE FROM N.A.
RA Korkko J.M., Earley J.J., Ala-Kokko L., Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by CSGE and DNA sequencing in
RT 14 patients with mild OI (Type I). Identification of common sequences
RT for null allele mutations.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88339824; PubMed=3421913;
RA Kuivaniemi H., Tromp G., Chu M.-L., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 2(I)
RT chain of human type I procollagen. Comparison with the chicken gene
RT confirms unusual patterns of gene conservation.";
RL Biochem. J. 252:633-640(1988).
RN [4]
RP SEQUENCE OF 181-1366 FROM N.A.
RA Kalicki J., Wamsley P., Gibson A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 623-1366 FROM N.A.
RA Bernard M.P., Myers J.C., Chu M.-L., Ramirez F., Eikenberry E.F.,
RT Prockop D.J.;
RT "Structure of a cDNA for the pro alpha 2 chain of human type I
RT procollagen. Comparison with chick cDNA for pro alpha 2(I) identifies
RT structurally conserved features of the protein and the gene.";
RL Biochemistry 22:1139-1145(1983).
RN [6]
RP SEQUENCE OF 80-96.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [7]
RP SEQUENCE OF 417-447.
RC TISSUE=Skin;
RX MEDLINE=75008198; PubMed=4412529;
RA Fietzek P.P., Furtmayr H., Kuehn K.;
RT "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
RT and pig-skin collagen.";
RL Eur. J. Biochem. 47:257-261(1974).
RN [8]
RP SEQUENCE OF 145-198 FROM N.A.
RX MEDLINE=88298792; PubMed=3403536;
RA Kuivaniemi H., Sabol C., Tromp G., Sippola-Thiele M., Prockop D.J.;
RT "A 19-base pair deletion in the pro-alpha 2(I) gene of type I
RT procollagen that causes in-frame RNA splicing from exon 10 to exon 12
RT in a proband with atypical osteogenesis imperfecta and in his
RT asymptomatic mother.";
RL J. Biol. Chem. 263:11407-11413(1988).
RN [9]
RP SEQUENCE OF 960-1351 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=90304220; PubMed=2364107;
RA Maekelae J.K., Vuorio T., Vuorio E.;
RT "Growth-dependent modulation of type I collagen production and mRNA
RT levels in cultured human skin fibroblasts.";
RL Biochim. Biophys. Acta 1049:171-176(1990).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-

RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalgleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT EDS-VII-A2
RX MEDLINE=88059013; PubMed=3680255;
RA Wirtz M.K., Glanville R.W., Steinmann B., Rao V.H., Hollister D.W.;
RT "Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids
RT comprising the N-telopeptide region of a pro-alpha 2(I) chain.";
RL J. Biol. Chem. 262:16376-16385(1987).
RN [15]
RP SEQUENCE OF 1090-1107 FROM N.A., AND VARIANT OI-IV ARG-1102.
RX MEDLINE=88227975; PubMed=2897363;
RA Wenstrup R.J., Cohn D.H., Cohen T., Byers P.H.;
RT "Arginine for glycine substitution in the triple-helical domain of
RT the products of one alpha 2(I) collagen allele (COL1A2) produces the
RT osteogenesis imperfecta type IV phenotype.";
RL J. Biol. Chem. 263:7734-7740(1988).
RN [16]
RP VARIANT OI-II ASP-997.
RX MEDLINE=89123407; PubMed=2914942;
RA Baldwin C.T., Constantinou C., Dumars K.W., Prockop D.J.;
RT "A single base mutation that converts glycine 907 of the alpha 2(I)
RT chain of type I procollagen to aspartate in a lethal variant of
RT osteogenesis imperfecta. The single amino acid substitution near the
RT carboxyl terminus destabilizes the whole triple helix.";
RL J. Biol. Chem. 264:3002-3006(1989).
RN [17]
RP VARIANT OI-II SER-955.
RX MEDLINE=89380165; PubMed=2777764;
RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
RN [18]
RP VARIANT OI-II CYS-877.
RX MEDLINE=90368825; PubMed=2394758;
RA Weill D., D'Alessio M., Ramirez F., Eyre D.R.;
RT "Structural and functional characterization of a splicing mutation in
RT the pro-alpha 2(I) collagen gene of an Ehlers-Danlos type VII
RT patient.";
RL J. Biol. Chem. 265:16007-16011(1990).
RN [20]
RP VARIANTS OI-IV VAL-676.
RX MEDLINE=91291136; PubMed=2064612;
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN [21]
RP VARIANTS OI CYS-349 AND CYS-736.
RX MEDLINE=91115889; PubMed=1990009;

RA Wenstrup R.J., Shrago-Howe A.W., Lever L.W., Phillips C.L.,
 RA Byers P.H., Cohn D.H.;
 RT "The effects of different cysteine for glycine substitutions within the
 RT alpha 2(I) chains. Evidence of distinct structural domains within the
 RT type I collagen triple helix.";
 RL J. Biol. Chem. 266:2590-2594(1991).
 RN [22]
 RP VARIANT OI-II ARG-784.
 RX MEDLINE=91340689; PubMed=1874719;
 RA Tsuneyoshi T., Westerhausen A., Constantinou C.D., Prockop D.J.;
 RT "Substitutions for glycine alpha 1-637 and glycine alpha 2-694 of
 RT type I procollagen in lethal osteogenesis imperfecta. The
 RT conformational strain on the triple helix introduced by a glycine
 RT substitution can be transmitted along the helix.";
 RL J. Biol. Chem. 266:15608-15613(1991).
 RN [23]
 RP VARIANT OI-IV SER-751.
 RX MEDLINE=91271401; PubMed=2052622;
 RA Spotila L.D., Constantinou C.D., Sereda L., Ganguly A., Riggs B.L.,
 RA Prockop D.J.;
 RT "Mutation in a gene for type I procollagen (COL1A2) in a woman with
 RT postmenopausal osteoporosis: evidence for phenotypic and genotypic
 RT overlap with mild osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5423-5427(1991).
 RN [24]
 RP VARIANT OI-II ARG-547.
 RX MEDLINE=93244832; PubMed=1284475;
 RA Bateman J.F., Moeller I., Hannagan M., Chan D., Cole W.G.;
 RT "Lethal perinatal osteogenesis imperfecta due to a type I collagen
 RT alpha 2(I) Gly to Arg substitution detected by chemical cleavage of
 RT an mRNA: cDNA sequence mismatch.";
 RL Hum. Mutat. 1:55-62(1992).
 RN [25]
 RP VARIANT OI-II ASP-670.
 RX MEDLINE=93054637; PubMed=1385413;
 RA Query Match 28.1%; Score 48; DB 1; Length 1366;
 RA Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 RA Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;
 QY 1 IEGPTLRQWLAARAGPNIQGP 22
 Db : ||| | ||||| ||
 752 VVGPTGPGVGAAGPAGPAGP 773
 RESULT 23
 ID SYK_AERPE STANDARD; PRT; 562 AA.
 AC Q9VFT9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) (LYSRS).
 GN LYSS OR APE0161.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KJ;
 RX MEDLINE=93310339; PubMed=10382966;
 RA Kavarabasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 CC -I- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
 CC + L-lysyl-Lys(Lys).

CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -I- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP000058; BAA79072.1; -
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002904; tRNA-synt_lys_1.
 DR Pfam; PF01921; tRNA-synt_1f; 1.
 DR PROSITE; PS00178; AA-TRNA_LIGASE_I; FALSE-NEG.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 50 58 "HIGH" REGION.
 FT SITE 305 309 "KMSKS" REGION.
 SQ SEQUENCE 562 AA; 65114 MW; 753664E2937FBF27 CRC64;
 Query Match 27.8%; Score 47.5; DB 1; Length 562;
 Best Local Similarity 35.7%; Pred. No. 61;
 Matches 10; Conservative 5; Mismatches 6; Indels 7; Gaps 1;
 QY 8 QWLAARAG-----PNIQPTLRQWL 28
 Db : ||: ||| : | | | | |
 293 EWSLRAGGREADSMSSGFTGITPREWL 320
 RESULT 24
 ID FRT2_HUMAN STANDARD; PRT; 113 AA.
 AC O75474;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE GSK-3 binding protein FRAT2 (fragment).
 GN FRAT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98297355; PubMed=9635432;
 RA Yost C., Farr G.H. III, Pierce S.B., Ferkey D.M., Chen M.M.,
 RA Kimelman D.;
 RT "GBP, an inhibitor of GSK-3, is implicated in Xenopus development and
 RT oncogenesis.";
 RL Cell 93:1031-1041(1998).
 CC -I- FUNCTION: BINDS GSK-3 AND PREVENTS GSK-3-DEPENDENT
 CC PHOSPHORYLATION. MAY BE IMPLICATED IN TUMOR PROGRESSION.
 CC -I- SIMILARITY: BELONGS TO THE GSK-3-BINDING PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF062739; AAC39786.1; -
 DR MIM; 605006; -
 FT NON_TER 1 1
 FT DOMAIN 54 76 INVOLVED IN GSK-3 BINDING (BY
 FT SIMILARITY).
 SQ SEQUENCE 113 AA; 11779 MW; CCEC4E7746694AC CRC64;
 Query Match 27.5%; Score 47; DB 1; Length 113;
 Best Local Similarity 53.3%; Pred. No. 14;

```

KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
KW Complete proteome. 294 FMN (POTENTIAL);
NP_BIND 286
FT SEQUENCE 357 AA; 37998 MW; 3D9D107DD9B4FCB6 CRC64;
SQ

Query Match 27.5%; Score 47; DB 1; Length 357;
Best Local Similarity 64.3%; Pred. NO. 45;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 13 RAGPNGIEGPTLRQ 26
      | | | | | | | |
Db 256 RLGGPGGISGPPLAQ 269

RESULT 26
PSBC_PINTH STANDARD; PRT; 473 AA.
ID PSBC_PINTH
AC P41643;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
GN PSBC.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugira M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -!- FUNCTION: THE 43 kDa PROTEIN (P6) IS A COMPONENT OF THE CORE OF
CC PHOTOSYSTEM II. IT IS A CHLOROPHYLL BINDING PROTEIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; CHLOROPLAST
CC THYLAKOID MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D17510; BAA04424.1; -
CC Mendel; 10000; PINTH:psbc1.1.
CC InterPro; IPR000932; PSII.
CC Pfam; PF00421; PSII; 1.
KW Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Chloroplast;
KW Transmembrane.
SQ SEQUENCE 473 AA; 51826 MW; 0E05E8FC7268465C CRC64;

Query Match 27.5%; Score 47; DB 1; Length 473;
Best Local Similarity 32.5%; Pred. NO. 59;
Matches 13; Conservative 5; Mismatches 12; Indels 10; Gaps 2

QY 3 GPTLR-----OWLAARAGPNGIEGPTLRQ-----WLAARA 32
      | | | | | | | | | | | | | | | |
Db 353 GETMRFWDLRAPWLEPLRGPNGLDLKLRKDIQPWQERRS 392

RESULT 27
PUR4_HUMAN
ID PUR4_HUMAN STANDARD; PRT; 1338 AA.
AC O15067;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 11.466 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-22
Perfect score: 171
Sequence: 1 IEPTLTROWLAARAGPNEGPTLRLQWLAARA 32

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	63	36.8	683	16	O83436 treponema p
2	62	36.3	607	2	Q9L8D4
3	60	35.1	509	2	Q9S5E5
4	58.5	34.2	869	5	Q9VZ82
5	56	32.7	361	16	Q9ABC7
6	55.5	32.5	1744	3	O94192
7	55	32.2	420	2	P97011
8	54	31.6	1095	16	Q9I304
9	53	31.0	305	2	Q9S0M9
10	53	31.0	326	16	Q9RTB6
11	52.5	30.7	814	4	Q96C78
12	52	30.4	396	2	Q9X7N5
13	52	30.4	902	5	O36161
14	52	30.4	967	2	Q9KZD5
15	52	30.4	1349	2	Q9L096
16	51.5	30.1	371	16	Q9I477

17	51	29.8	281	17	Q9YDQ0
18	51	29.8	306	16	O05576
19	51	29.8	322	2	Q9RX51
20	51	29.8	381	2	Q9X757
21	51	29.8	589	5	Q18756
22	51	29.8	600	16	Q9HYJ8
23	51	29.8	719	16	Q922H9
24	51	29.8	1820	13	Q9I907
25	50.5	29.5	526	16	Q981N1
26	50.5	29.5	1460	5	Q9GY79
27	50	29.2	250	2	Q93LY8
28	50	29.2	268	16	Q98LG1
29	50	29.2	351	16	Q9RWB0
30	50	29.2	384	2	Q9F2F9
31	50	29.2	400	2	Q9XDB0
32	50	29.2	403	2	Q9F5B8
33	50	29.2	410	16	Q9IIM2
34	50	29.2	472	5	O17754
35	50	29.2	604	16	Q98P10
36	50	29.2	1272	4	Q9UGH1
37	50	29.2	1300	4	Q9BXA9
38	49.5	28.9	333	10	Q94LX0
39	49.5	28.9	336	11	Q9WV74
40	49	28.7	133	2	Q9AQH5
41	49	28.7	214	5	Q20968
42	49	28.7	249	2	Q9L3H3
43	49	28.7	250	10	Q9AS26
44	49	28.7	307	10	Q43416
45	49	28.7	319	2	Q9RRK5

ALIGNMENTS

RESULT 1

O83436 ID O83436 PRELIMINARY; PRT; 683 AA.
AC O83436;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0421.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiaach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001220; AAC65409.1; -
DR TIGR; TP0421; -
DR InterPro; IPR001258; NHL.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01436; NHL; 4.
DR Pfam; PF00515; TPR; 1.
KW Complete proteome.
SQ SEQUENCE 683 AA; 74518 MW; F91407FA7094AAD1 CRC64;

Query Match 36.8%; Score 63; DB 16; Length 683;
Best Local Similarity 46.4%; Pred. No. 4.6;
Matches 13; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Q9YDQ0 aeropyrum p
O05576 mycobacteri
Q9RX51 streptomyce
Q9X757 klebsiella
Q18756 caenorhabdi
Q9HYJ8 pseudomonas
Q922H9 rhizobium m
Q9I907 pagrus major
Q981N1 rhizobium l
Q9GY79 leishmania
Q93LY8 streptomyce
Q98LG1 rhizobium l
Q9RWB0 deinococcus
Q9F2F9 streptomyce
Q9XDB0 mycobacteri
Q9F5B8 agrobacteri
Q9IIM2 pseudomonas
O17754 caenorhabdi
Q98P10 rhizobium l
Q9UGH1 homo sapien
Q9BXA9 homo sapien
Q94LX0 perilla fru
Q9WV74 mus musculu
Q9AQH5 achromobact
Q20968 caenorhabdi
Q9L3H3 rhizobium l
Q9AS26 oryza sativ
Q43416 cenchrus ci
Q9RKM5 streptomyce

```
QY 4 PTLROWLAARAGPNGIEGPTLROWLAAR 31
  1 : : : : : : : : : : : : : : : :
Db 74 PTLWLGNAYRSGIEGAALHGWGAAR 101

RESULT 2
Q9L8D4
ID Q9L8D4 PRELIMINARY; PRT; 607 AA.
AC Q9L8D4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 66.3 KDA PROTEIN (FRAGMENT).
OS Polyangium cellulosum.
OC Bacteria; proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SO CE90;
RX MEDLINE=20130945; PubMed=10662695;
RA Molnar I., Schupp T., Ono M., Zickler R.E., Milnamow M.,
RA Nowak-Thompson B., Engel N., Toupet C., Stratmann A., Cyr D.D.,
RA Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT "The biosynthetic gene cluster for the microtubule-stabilizing agents
RL epothilones A and B from Sorangium cellulosum So ce90.";
RL Chem. Biol. 7:97-109(2000).
DR EMBL; AF210843; AAF26904.1; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 607 AA; 66326 MW; F113CA299B25048E CRC64;
  Query Match 36.3%; Score 62; DB 2; Length 607;
  Best Local Similarity 34.8%; Pred. No. 5.5;
  Matches 16; Conservative 3; Mismatches 7; Indels 20; Gaps 2;

QY 1 IEGPTLROWLAARAGPNGIEGPTLROWLAAR 30
  1 : : : : : : : : : : : : : : : :
Db 404 LAGPALRTWAVDLLGRLDPDGRDLRRLRTLRTWIAA 437

RESULT 4
Q9VZ82
ID Q9VZ82 PRELIMINARY; PRT; 869 AA.
AC Q9VZ82;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG7479 PROTEIN.
DE CG7479.
GN CG7479.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
```

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003482; AAF47943.1; -.
DR FlyBase; FBgn0035576; CG7479.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002302; tRNA-synt_leu.
DR Pfam; PF00133; tRNA-syntL1; 1.
DR PRINTS; PR00985; TRNASYNTHLEU.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ SEQUENCE 869 AA; 99299 MW; EB7A1ECBEBB27B67 CRC64;

Query Match 34.2%; Score 58.5; DB 5; Length 869;
Best Local Similarity 40.6%; Pred. No. 22;
Matches 13; Conservative 4; Mismatches 10; Indels 5; Gaps 1;

QY 1 IEPTLRQWLA-----ARAGPNNGIEGPTLRQW 27
DB 213 VEKLLRQWFTRTSAYAKQLDGLGEDPTLRDW 244

RESULT 5
Q9ABC7 PRELIMINARY; PRT; 361 AA.
AC Q9ABC7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CATION EFFLUX FAMILY PROTEIN.
GN CC0303.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005704; AAK22290.1; -.
DR TIGR; CC0303; -.
DR InterPro; IPR002524; Cation_efflux.
DR InterPro; IPR002395; Kininogen.
DR Pfam; PF01545; Cation_efflux; 1.
DR PRINTS; PR00334; KININOGEN.
KW Complete proteome.
SQ SEQUENCE 361 AA; 38180 MW; 1A4F7FOA7C62EBE0 CRC64;

Query Match 32.7%; Score 56; DB 16; Length 361;
Best Local Similarity 54.5%; Pred. No. 19;
Matches 12; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 10 LAARAGPNNGIEGPTLRQWLAAR 31
DB 266 LALDATPRGIDTQKVRDWLAAR 287

RESULT 6
O94192 PRELIMINARY; PRT; 1744 AA.
AC O94192;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

```

```

DE CHITIN SYNTHASE.
GN CHS4.
OS Paracoccidioides brasiliensis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Onygenales; mitosporic Onygenales; Paracoccidioides.
OX NCBI_TaxID=121759;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=20210320; PubMed=10746225;
RA Nino-Vega G.A., Munro C.A., San-Blas G., Gooday G.W., Gow N.A.;
RT "Differential expression of chitin synthase genes during temperature-
RT induced dimorphic transitions in Paracoccidioides brasiliensis.";
RL Med. Mycol. 38:31-39(2000).
RN 2
RP SEQUENCE FROM N.A.
RA Nino-Vega G.A., San-Blas G.;
RT "Sequence analysis of the CHS4 gene of Paracoccidioides
RT brasiliensis.";
RL EMBL: AF107624; RAD19613.2; -.
DR InterPro; IPR002923; Chitin_synth.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR001173; Glycos_transf_2.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF03142; Chitin_synth_2; 1.
DR Pfam; PF00063; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
SQ SEQUENCE 1744 AA; 193777 MW; DB7622D0A69F0705 CRC64;

Query Match 32.5%; Score 55.5; DB 3; Length 1744;
Best Local Similarity 51.7%; Pred. No. 11e+02;
Matches 15; Conservative 2; Mismatches 11; Indels 1; Gaps 1;

QY 5 TLRQWL-AARAGPNNGIEGPTLRQWLAAR 32
DB 56 TVNTWLTAASPGNGEVGGTIDADLARRA 84

RESULT 7
ID P97011 PRELIMINARY; PRT; 420 AA.
AC P97011;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SORBITOL OXIDASE.
GN SOX.
OS Streptomyces sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1931;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=H-7775;
RA Hiraga K., Eto T., Yoshioka I., Oda K.;
RT "Cloning of a gene encoding a sorbitol oxidase from Streptomyces sp.
RT H-7775.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB000519; BAAL9135.1; -.
DR InterPro; IPR001575; Oxid_FAD_bind.
DR Pfam; PF01565; FAD_binding_4; 1.
SQ SEQUENCE 420 AA; 45181 MW; EF3189045CAF0649 CRC64;

Query Match 32.2%; Score 55; DB 2; Length 420;
Best Local Similarity 37.9%; Pred. No. 29;
Matches 11; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 3 GPTLRQWLAARAGPNNGIEGPTLRQWLAAR 31
DB 215 GFVQVWLKQVRGDEGARSVMPAEWLGAR 243

```

RESULT 8					
D	C	Q913U4	PRELIMINARY;	PRT:	1095 AA.
C	C	Q913U4:			
O1-MAR-2001	(TREMBlrel.. 16, Created)				
O1-MAY-2000	(TREMBlrel.. 16, Last sequence update)				
O1-DEC-2001	(TREMBlrel.. 19, Last annotation update)				
E	E	PROBABLE PYRUVATE CARBOXYLASE.			
N	N	PA1400.			
S	S	Pseudomonas aeruginosa.			
C	C	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
C	C	Pseudomonas.			
C	C	NCBI_TaxID=287;			
R	R	[1]			
C	C	SEQUENCE FROM N.A.			
X	X	STRAIN=ATCC 15692 / PAO1;			
X	X	MEDLINE=20437337; PubMed=10984043;			
A	A	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,			
A	A	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,			
A	A	Garber R.L., Goltzman L., Tolentino E., Westbrook-Wadman S., Yuan Y.,			
A	A	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,			
A	A	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
A	A	Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;			
R	R	*Complete genome sequence of Pseudomonas aeruginosa PAO1, an			
T	T	opportunistic pathogen.*;			
L	L	Nature 406:959-964(2000).			
C	C	-I- COFACTOR: BIOTIN (BY SIMILARITY).			
R	R	EMBL: AE004569; AAG04789.1; -.			
R	R	HSP: P24182; lBNC.			
R	R	InterPro: IPR001249; AcCoA_biotinCC.			
R	R	InterPro: IPR001892; Biotin.			
R	R	InterPro: IPR000089; Biotin_lipoyl.			
R	R	InterPro: IPR000022; Carboxyl_trans.			
R	R	InterPro: IPR000901; CPase.			
R	R	InterPro: IPR001064; Crystallin.			
R	R	pfam: PF02785; Biotin_carb_C; 1.			
R	R	pfam: PF00364; biotin_lipoyl; 1.			
R	R	pfam: PF01039; Carboxyl_trans; 1.			
R	R	pfam: PF00289; CPase_L_chain; 1.			
R	R	pfam: PF02786; CPase_L_D2; 1.			
R	R	PROSITE: PS00188; BIOTIN; 1.			
R	R	PROSITE: PS00867; CPASE_2; UNKNOWN_1.			
R	R	PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOW_N_1.			
KW	KW	Biotin: Complete proteome; Pyruvate.			
SSQ	SSQ	SEQUENCE 1095 AA; 116876 MW; 34370FB8BEC201AD CRC64;			
QY	QY	Query Match 31.6%; Score 54; DB 16; Length 1095;			
DB	DB	Best Local Similarity 45.5%; Pred.No.1.1e-02;			
		Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;			
QY	QY	1 TEGPTLRQLRWLAARAGPNGIEGP 22			
DB	DB	786 LEGGLGRFAAEVGTGVQGP 807			
RESULT 9					
Q9SOM9	Q9SOM9	PRELIMINARY;	PRT:	305 AA.	
ID	ID	Q9SOM9			
AC	AC	Q9SOM9;			
DT	DT	O1-MAY-2000 (TREMBlrel.. 13, Created)			
TT	TT	O1-MAY-2000 (TREMBlrel.. 13, Last sequence update)			
DT	DT	O1-MAY-2000 (TREMBlrel.. 13, Last annotation update)			
DE	DE	UV-ENDONUCLEASE.			
GN	GN	UVSCDE.			
OS	OS	Deinococcus radiodurans.			
OC	OC	Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.			
OX	OX	NCBI_TaxID=1299;			
RN	RN	[1]			
R	R	SEQUENCE FROM N.A.			
RC	RC	STRAIN=KRL;			
RA	RA	Kitayama S., Kikuchi M., Funayama T., Narumi I., Watanabe H.;			
RT	RT	"Cloning of structural gene of an alternative incision enzyme for DNA			
RP	RP	damage in Deinococcus radiodurans.";			


```

SQ  SEQUENCE      814 AA:  87717 MW:  683A8368AD30996B  CRC64;

Query Match          30.7%; Score 52.5; DB 4; Length 814;
Best Local Similarity 44.8%; Pred. No. 1.2e+02;
Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY  1 IEQPTLRQWLAAAGAGNGIEGPTLRQWLA 29
      ::||| 1: ||::||: || | ||
Db   728 LKGPCTC-QYRAAGSGSERPGPPQRAALLA 755

RESULT 12
Q9X7N5
ID   Q9X7N5          PRELIMINARY;      PRT;       396 AA.
AC   Q9X7N5;
DT   01-NOV-1999 (TREMBLrel. 12, Created)
DT   01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   SC5FZA.12C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Oliver K., Harris D.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL049587; CAB40679.1; -.
KW   Hypothetical protein.
SQ  SEQUENCE      396 AA:  41908 MW:  BCB465197F3A3F6E  CRC64;

Query Match          30.4%; Score 52; DB 2; Length 396;
Best Local Similarity 34.4%; Pred. No. 66;
Matches 11; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

QY  4 PTLRQWLAAAGAGNGIEGPT---LRQWLAAR 31
      ||::||| 1: |||: ||: ||: ||: ||: ||
Db   293 PPARRWLGRLAGP--EGPSAERRAKSWFSVR 322

RESULT 13
O16161
ID   O16161          PRELIMINARY;      PRT;       902 AA.
AC   O16161;
DT   01-JAN-1998 (TREMBLrel. 05, Created)
DT   01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   PRECOLLAGEN P PRECURSOR.
GN   PRECOL-P.
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=6550;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FOOT;
RX   MEDLINE=97442537; PubMed=9295257;

```

```

DR   InterPro; IPR000205; NAD_binding.
SQ   Hypothetical protein; Complete proteome.
KW   SEQUENCE 371 AA; 39174 MW; 016D50440BAD50D7 CRC64;

Query Match      30.1%; Score 51.5; DB 16; Length 371;
Best Local Similarity 31.0%; Pred. No. 72;
Matches 13; Conservative 7; Mismatches 9; Indels 13;

QY   4   PTLRWLAARAGP-----NGIEGPTLR-----QWLAARA 32
      |   :|| :||| :||| :||| :||| :||| :||| :||| :|||
DB   145 PNAARWLLDQAGPRLLRLRYAEVSEVDGSRKLRADGRWLSEA 186

RESULT 17
Q9YDQ0
ID   Q9YDQ0 PRELIMINARY; PRT; 281 AA.
AC   Q9YDQ0;
DT   01-NOV-1999 (TREMBLrel. 12, Created)
DT   01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 32.1 KDA PROTEIN APE0867.
GN   APE0867.
OC   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RS   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.
RA   Jin-n K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosi
RA   Hosoyana A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushiida N., Ouchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000060; BAA79847.1; -.
DR   InterPro; IPR000130; Zn_MTpeptidse.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 281 AA; 32123 MW; 09AC9AF6F92CB41E CRC64;

Query Match      29.8%; Score 51; DB 17; Length 281;
Best Local Similarity 34.4%; Pred. No. 62;
Matches 11; Conservative 7; Mismatches 6; Indels 8;

QY   5   TLRWLAARAGPN-----GIEGPTLRQWLAAR 31
      |||||: :|| :||| :||| :||| :||| :||| :||| :|||
DB   12 SLRWMRS---ENRYDIPVDSPVEGWWLESR 40

RESULT 18
O05576
ID   O05576 PRELIMINARY; PRT; 306 AA.
AC   O05576;
DT   01-JUL-1997 (TREMBLrel. 04, Created)
DT   01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   GALU.
GN   GALU OR RV0993 OR MTC1237.07.
OC   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobact
OX   NCBI_TaxID=1773;
RN   [1]
RS   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., He
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

```

AC	Q9X757;	
AD	01-NOV-1999 (TrEMBLrel. 12, Created)	
AE	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)	
AF	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
AG	BETA_LACTAMASE.	
AH	MIR-1.	
AI	Klebsiella pneumoniae.	
AJ	Plasmid pMG230.	
AK	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
AL	Klebsiella.	
AM	NCBI_TaxID=573;	
AN	[1]	
AO	SEQUENCE FROM N.A.	
AP	MEDLINE=91159299; PubMed=1963529;	
AQ	Papanicolaou G.A., Medeiros A.A., Jacoby G.A.;	
AR	"Novel plasmid mediated beta-lactamase (MIR-1) conferring resistance	
AS	to oxyimino- and alpha-methoxy-beta-lactams in clinical isolates of	
AT	Klebsiella pneumoniae.";	
AV	Antimicrob. Agents Chemother. 34:2200-2209(1990).	
AW	[2]	
AX	SEQUENCE FROM N.A.	
AY	Jacoby G.A., Tran J.;	
AZ	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.	
BA	EMBL; M37839; AAD22636.1; -.	
BB	HSP; P05364; 2BLT.	
BC	InterPro; IPR001466; Beta_Lactam.	
BD	InterPro; IPR001586; Beta_Lactam.C.	
BE	Pfam; PF00144; beta-lactamase; 1.	
BF	PROSITE; PS00336; BETA_LACTAMASE_C; 1.	
BG	Plasmid.	
BH	SEQUENCE - 381 AA; 41171 MW; DD581D789C03142E CRC64;	
BI	Query Match	29.88; Score 51; DB 2; Length 381;
BJ	Best Local Similarity	33.38; Pred.No. 85;
BK	Matches	8; Conservative 7; Mismatches 9; Indels 0; Gaps
BL		
BM		
BN		
BO		
BP		
BQ	6 LRQWLAARAGPNIGEGPTLRQWLA 29	
BR	: : : : : :	
BS	250 MASWLIANKPKPDSLQAPSLKOGIA 273	
BT		
BU		
BV		
BW		
BX		
BY		
BZ		
CA	RESULT 21	
CB	Q18756	
CC	Q18756 PRELIMINARY; PRT; 589 AA.	
CD	AC Q18756;	
CE	01-NOV-1996 (TrEMBLrel. 01, Created)	
CF	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)	
CG	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
CH	CS0F7.2 PROTEIN.	
CI	C50F7.2.	
CJ	GN Caenorhabditis elegans.	
CK	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;	
CL	Rhabditidae; Pelodirinae; Caenorhabditis.	
CM	NCBI_TaxID=6239;	
CN	[1]	
CO	SEQUENCE FROM N.A.	
CP	MEDLINE=94150718; PubMed=7906398;	
CQ	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,	
CR	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,	
CS	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,	
CT	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,	
CU	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,	
CV	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,	
CW	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,	
CX	Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,	
CY	Thierry-Mieg J., Thomas K., Vaughan M., Vaughan K., Waterston R.,	
CA	Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;	
CB	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.	
CC	elegans.";	
CD	Nature 368:32-38(1994).	
CE	[2]	
CF	SEQUENCE FROM N.A.	
CG	Johnson, D. Stellives L.;	
CH	RP	

Query Match 29.8%; Score 51; DB 13; Length 1820;

Best Local Similarity 52.6%; Pred. No. 4.5e+02;

Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 GPTLRQWLAARAGPNGTEG 21

DB 1403 GPVGPQGLAGKAGPEGLRG 1421

RESULT 25

ID Q981N1 PRELIMINARY; PRT; 526 AA.

AC Q981N1;

DT 01-OCT-2001 (TREMBLrel. 18, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)

DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE REPLICATION PRIMASES.

GN MLL9305.

OS Rhizobium loti (Mesorhizobium loti).

OG Plasmid pMLa.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Phyllobacteriaceae; Mesorhizobium.

OX NCBI_TaxID=381;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MAFF303099;

RX MEDLINE=21082930; PubMed=11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,

RA Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RL DNA Res. 7:331-338(2000).

DR EMBL: AP003015; BAB54678.1; -.

KW Plasmid; Complete proteome.

SQ SEQUENCE 526 AA; 57216 MW; A20B4A4F13C98BD7 CRC64;

Query Match

Best Local Similarity 50.0%; Score 50.5; DB 16; Length 526;

Matches 15; Conservative 3; Mismatches 9; Indels 3; Gaps 2;

QY 5 TLRLQWLAAR--AGPNGIEGTTLRLQWLAARA 32

DB 283 TLALWVAQAQAAPMGWLG-TYRQWQALGA 311

RESULT 26

ID Q9GY79 PRELIMINARY; PRT; 1460 AA.

AC Q9GY79;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE PROBABLE HYPOTHETICAL 21.3 KDA PROTEIN.

GN M12.50.

OS Leishmania major.

OC Eukaryota; Eulgenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI_TaxID=5664;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FRIDLIN;

RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,

RA Oliver K.;

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL390114; CAC01975.2; -.

DR InterPro: IPR001313; PUM.

DR Pfam: PF00806; PUF; 6.

DR SMART: SM00025; Pumilio; 7.

SQ SEQUENCE 1460 AA; 153236 MW; 7B5EF90010364DD0 CRC64;

Query Match

Best Local Similarity 29.5%; Score 50.5; DB 5; Length 1460;

Best Local Similarity 37.9%; Pred. No. 4.1e+02;

Matches 11; Conservative 6; Mismatches 7; Indels 5; Gaps 1;

QY 4 PTLRQWLAARAGP-----NGIEGPTLRQW 27

DB 725 PSLATAAAAAAGPYKQONHQTPTSMRW 753

RESULT 27

Q93LY8

ID Q93LY8 PRELIMINARY; PRT; 250 AA.

AC Q93LY8;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE PGAK (FRAGMENT).

GN PGAK.

OS Streptomyces sp. PGA64.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=161235;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PGA64;

RA Metsa-Ketela M., Kantola J., Ylihonko K.;

RT "Cloning and Characterization of a Silent Angucycline-type Gene

RT Cluster from a Rubromycin B Producing Streptomyces sp. PGA64.;"

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY034378; AAK57521.1; -.

FT NON_TER 250 250

SQ SEQUENCE 250 AA; 26031 MW; 597EB6581FF8C97 CRC64;

Query Match

Best Local Similarity 29.2%; Score 50; DB 2; Length 250;

Matches 15; Conservative 2; Mismatches 10; Indels 12; Gaps 2;

QY 6 LRQWLAARAGPNGIEGPT-----LR--QWLAARA 32

DB 110 LASNSAAVQAERGVAAPTRVVARLAGFLLRHIEWLAHA 148

RESULT 28

Q98LGI

ID Q98LGI PRELIMINARY; PRT; 268 AA.

AC Q98LGI;

DT 01-OCT-2001 (TREMBLrel. 18, Created)

DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)

DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE PROBABLE SHORT CHAIN DEHYDROGENASE.

GN MLL1036.

OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Phyllobacteriaceae; Mesorhizobium.

OX NCBI_TaxID=381;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MAFF303099;

RX MEDLINE=21082930; PubMed=11214968;

RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,

RA Takeuchi C., Yamada M., Tabata S.;

RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RL DNA Res. 7:331-338(2000).

DR EMBL: AP002996; BAB48502.1; -.

DR InterPro: IPR002198; ADH_Short.

DR InterPro: IPR000205; NAD_Binding.

DR Pfam: PF00106; adh_short; 1.

DR PRINTS: PR00080; SDRFAMILY.

DR Complete proteome.

KW SEQUENCE 268 AA; 27788 MW; 86698FFD04036653 CRC64;

Query Match 29.2%; Score 50; DB 16; Length 268;
Best Local Similarity 40.0%; Pred. No. 79;
Matches 14; Conservative 3; Mismatches 8; Indels 10; Gaps 1;

QY 8 QWLAARAGPNG-----TEGPTLRWLAAARA 32
| | | | | : | | | : | | | |
Db 176 OSLAKELGPHGIRVNAILPGIIEGPRIGVIAARA 210

RESULT 29

ID Q9RWB0 PRELIMINARY; PRT; 351 AA.
AC Q9RWB0;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN DR0759.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001931; AAF10338.1; -.
DR TIGR; DR0759; -.
KW Complete proteome.
SQ SEQUENCE 351 AA; 38454 MW; 31B01305A7B28694 CRC64;

Query Match 29.2%; Score 50; DB 16; Length 351;
Best Local Similarity 34.2%; Pred. No. 1.le+02;
Matches 13; Conservative 3; Mismatches 6; Indels 16; Gaps 1;

QY 8 QWLAARAGPNGIEGPTLRQ-----WLA 29
| | | | | : | | | : | | | |
Db 206 QGIADRFPHRIDGPDYRQRTGTEPAQPLSEAEFAAWLA 243

RESULT 30

ID Q9F2F9 PRELIMINARY; PRT; 384 AA.
AC Q9F2F9;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE ELLORAMYCIN GLYCOSYLTRANSFERASE.
GN ELMGT.
OS Streptomyces olivaceus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=47716;
RN [1]
RP SEQUENCE FROM N.A.
RA Blanco G., Patallo E.P., Brana A.F., Trefzer A., Bechthold A.,
RA Rohr J., Mendez C., Salas J.A.;
RT "Identification of a sugar flexible glycosyltransferase from
RT Streptomyces olivaceus, the producer of the antitumor polyketide
RT elloramycin.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RA Patallo E.P.;
RT "Deoxysugar methylation during biosynthesis of the antitumor
RT polyketide elloramycin by Streptomyces olivaceus: characterization of
RT three methyltransferase genes.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Blanco G.;
RT "Identification of a sugar flexible glycosyltransferase from
RT Streptomyces olivaceus, the producer of the antitumor polyketide
RT elloramycin.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ300305; CAC16413.1; -.
DR EMBL; AJ309821; CAC32467.1; -.
DR InterPro; IPR000890; Acetate_kin.
DR PROSITE; PS01076; ACETATE_KINASE_2; UNKNOWN_1.
KW Transferase.
SQ SEQUENCE 384 AA; 39674 MW; A254F56B6ED12F2B CRC64;

Query Match 29.2%; Score 50; DB 2; Length 384;
Best Local Similarity 52.2%; Pred. No. 1.2e+02;
Matches 12; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 6 LRQWLAARAGPNGIEGPTLRQWL 28
| | | | | : | | | : | | | |
Db 152 VRQLAERLGPAGSEPPPERYFL 174

Search completed: October 9, 2002, 09:03:02
Job time : 13.5494 secs

Result No.	Score	Query %		Length	DB	ID	Description
		Match	Match				
1	171	100.0	32	21	AAB17297	TPO-mimetic peptid	
2	171	100.0	32	21	AAV96520	Thrombopoietin mim	
3	171	100.0	34	21	AAV96527	Thrombopoietin mim	
4	156	91.2	32	21	AAB17289	TPO-mimetic peptid	
5	147.5	85.3	31	21	AAB17288	TPO-mimetic peptid	
6	147	86.0	30	21	AAB17287	TPO-mimetic peptid	
7	145.5	85.1	33	21	AAB17290	TPO-mimetic peptid	
8	145	84.8	34	21	AAB17291	TPO-mimetic peptid	
9	145	84.8	36	21	AAB17306	TPO-mimetic peptid	
10	145	84.8	36	21	AAV96526	Thrombopoietin mim	
11	144.5	84.5	35	21	AAB17292	TPO-mimetic peptid	

ALIGNMENTS

RESULT 1

AAB17297	AAB17297 standard; Peptide; 32 AA.
ID	AAB17297 standard; Peptide; 32 AA.

AC AAB17297:

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:353.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer; autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF; immunosuppressive; EPO; TPO; CPLA; mimetic; IL-1; TNF; antagonist; MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1; cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor; vascular endothelial growth factor; matrix metalloproteinase; asthma; thrombosis; pharmaceutical.

Synthetic.

WO200024782-A2.

04-MAY-2000.

25-OCT-1999; 99WO-US25044 -

23-OCT-1998; 98US-0105371.

001 1999, 9303 0420002.

(FIFTH) FIFTEEN INC.

— 2250 м, снежный покров 5, воле 10;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 320; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-p1, -(L1)c-p1-(L2)d-p2,
CC -(L1)c-p1-(L2)d-p2-(L3)e-p3, or -(L1)c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4
CC where F1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antitumor, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present inventions can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAB69443
CC to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 32 AA;

Query Match 100.0%; Score 171; DB 21; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.8e-17;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEPTLRLQWLAARAGPNEGPTLRLQWLAARA 32
Db 1 IEPTLRLQWLAARAGPNEGPTLRLQWLAARA 32

RESULT 2
AAY96520
ID AAY96520 standard; peptide; 32 AA.
AC AAY96520;
XX
XX
XX 04-SEP-2000 (first entry)
XX Thrombopoietin mimetic peptide compound 1.
DE Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "optionally linked to an Fc molecule"
FT Peptide 1..14 /label= TMP_1
FT Peptide 15..18 /label= linker
FT Peptide 19..32 /label= TMP_2
FT Modified-site 32 /note= "optionally linked to an Fc molecule"
FT
FT
XX WO200024770-A2.
XX
XX 04-MAY-2000.
XX 22-OCT-1999; 99WO-US24834.
XX 23-OCT-1998; 98US-0105348.
XX

PA (AMGE-) AMGEN INC.
XX
XX Liu C, Feige U, Cheatham J;
XX DR WPI; 2000-365108/31.
XX
XX Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thrombocytopenia
XX
XX Claim 16; Page 61; 91pp; English.
XX
XX A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2].
CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X_2-X_1-0, X_2-X_1-1, X_2-X_1-2,
CC X_2-X_1-3, X_2-X_1-4, X_1-X_1-0, X_1-X_1-1, X_1-X_1-2, X_1-X_1-3, and
CC X_1-X_1-4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
CC or E; X_9 = W, Y or F; X_10 = L, I, V, A, F, M, or K; X_11 = A, I, V,
CC L, F, S, T, K, H, or G; X_12 = A, I, V, L, F, T, R, E, or G; L_1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
SQ Sequence 32 AA;

Query Match 100.0%; Score 171; DB 21; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.8e-17;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IEPTLRLQWLAARAGPNEGPTLRLQWLAARA 32
Db 1 IEPTLRLQWLAARAGPNEGPTLRLQWLAARA 32

RESULT 3
AAY96527
ID AAY96527 standard; peptide; 34 AA.
XX
XX
XX AAY96527;
XX
XX 04-SEP-2000 (first entry)
XX Thrombopoietin mimetic peptide compound 8.
DE Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "optionally linked to an Fc molecule"
FT Peptide 3..16 /label= TMP_1
FT Peptide 17..20 /label= linker
FT Peptide 21..34 /label= TMP_2
FT
FT
XX WO200024770-A2.
XX
XX 04-MAY-2000.
XX 22-OCT-1999; 99WO-US24834.
XX

Wed Oct 9 10:29:31 2002

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

XX
SQ Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;
Best Local Similarity 83.3%; Pred. NO. 1.4e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 IEGPTLRQWLAARA----GPNIGIEGPTLRQWLAARA 32
|||||
Db 1 IEGPTLRQWLAARAAGGGGGGIEGPTLRQWLAARA 36
|||||

RESULT 14
AAB17301
ID AAB17301 standard; Peptide; 36 AA.
XX
AC AAB17301;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:357.
XX
KW Modified peptide: therapeutic agent; fusion; Fc domain; cancer; autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF; immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist; MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1; cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor; vascular endothelial growth factor; matrix metalloproteinase; asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
XX
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -
XX
PS Example 1; Page 321; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

XX
SQ Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;
Best Local Similarity 83.3%; Pred. NO. 1.4e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 IEGPTLRQWLAARA----GPNIGIEGPTLRQWLAARA 32
|||||
Db 1 IEGPTLRQWLAARAAGGGGGGIEGPTLRQWLAARA 36
|||||

RESULT 15
AAB17303
ID AAB17303 standard; Peptide; 36 AA.
XX
AC AAB17303;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:359.
XX
KW Modified peptide: therapeutic agent; fusion; Fc domain; cancer; autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF; immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist; MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1; cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor; vascular endothelial growth factor; matrix metalloproteinase; asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
XX
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -
XX
PS Example 1; Page 322; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions can
 CC be useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 IEGPTLRQWLAAARA----GPNIGIEGPTLRQWLAAARA 32
 |||||
 Db 1 IEGPTLRQWLAAARAGCGGGGIEGPTLRQWLAAARA 36

RESULT 16

AAB17307
 ID AAB17307 standard; Peptide: 36 AA.

XX AAB17307;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:363.

XX Modified peptide: therapeutic agent; fusion: Fc domain; cancer;
 XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 324; 608pp; English.

XX The present invention describes composition of matter (1) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (1) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 IEGPTLRQWLAAARA----GPNIGIEGPTLRQWLAAARA 32
 |||||
 Db 1 IEGPTLRQWLAAARAGCGGGGIEGPTLRQWLAAARA 36

RESULT 17

AAAY96523
 ID AAY96523 standard; peptide: 36 AA.

XX AAY96523;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 4.

XX Thrombopoietin; mimetic; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
 XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14

FT Peptide /label= TMP_1

FT Peptide 15..22

FT Peptide /label= linker

FT Modified-site 18

FT Peptide /note= "optionally modified by bromoacetyl or PEG"

FT Peptide 23..36

FT Peptide /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,

CC X2-X13, X2-X14, X1-X10, X1-X11, X1-X12, X1-X13, and
 CC X1-X14, X1-I, A, V, L, S or R; X2-E, D, K or V; X3-G or A;
 CC X4-P; X5-T or S; X6-L, I, V, A or F; X7-R or K; X8-Q, N,
 CC or E; X9-W, Y or F; X10-L, I, V, A or F; X11-A, I, V,
 CC L, F, S, T, K, H, or E; X12-A, I, V, L, F, G, S, or Q; X13-R, K,
 CC T, V, N, Q or G; X14-A, I, V, L, F, T, R, E, or G; L1-linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 36 AA;
 Query Match 84.2%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. NO. 1.4e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAARA---GPNIEGPTLRQWLAARA 32
 |||||
 Db 1 IEPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 18
 AAY96524
 ID AAY96524 standard; peptide; 36 AA.

XX AC AAY96524;

XX DT 04-SEP-2000 (first entry)

XX DE Thrombopoietin mimetic peptide compound 5.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX OS Synthetic.

XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14 /label= TMP_1

FT Disulfide-bond 9..31 /note= "optional"

FT Peptide 15..22 /label= linker

FT Peptide 23..36 /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Claim 16; Page 62; 91pp; English.

CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1(L1)_TMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X2-X10, X2-X11, X2-X12,
 CC X2-X13, X2-X14, X1-X10, X1-X11, X1-X12, X1-X13, and
 CC X1-X14. X1-I, A, V, L, S or R; X2-E, D, K or V; X3-G or A;
 CC X4-P; X5-T or S; X6-L, I, V, A or F; X7-R or K; X8-Q, N,
 CC or E; X9-W, Y or F; X10-L, I, V, A, F, M, or K; X11-A, I, V,
 CC L, F, S, T, K, H, or E; X12-A, I, V, L, F, G, S, or Q; X13-R, K,
 CC T, V, N, Q or G; X14-A, I, V, L, F, T, R, E, or G; L1-linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 1.4e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAARA---GPNIEGPTLRQWLAARA 32
 |||||
 Db 1 IEPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 19
 AAY96525

ID AAY96525 standard; peptide; 36 AA.

XX AC AAY96525;

XX DT 04-SEP-2000 (first entry)

XX DE Thrombopoietin mimetic peptide compound 6.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.

XX OS Synthetic.

XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14 /label= TMP_1

FT Peptide 15..18 /label= linker

FT Peptide 19..32 /label= TMP_2

FT Modified-site 32 /note= "optionally linked to an Fc molecule"

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of

PT diseases which involve thrombocytopenia

PS Claim 16; Page 62; 9lpp: English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-TMP-2],
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
 CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A, F, M, or K; X-1-1 = A, I, V,
 CC L, F, S, T, K, H, or E; X-1-2 = A, I, V, L, F, G, S, or Q; X-1-3 = R, K,
 CC T, V, N, Q or G; X-1-4 = A, I, V, L, F, T, R, E, or G; L-1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

Query Match 84.2%; Score 144; DB 21; Length 36;

Best Local Similarity 83.3%; Pred. No. 1.4e-13;

Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAARA---GPNIEGPTLRQWLAARA 32

DB 1 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 36

RESULT 20

AAAY96528

ID AAAY96528 standard; peptide; 41 AA.

XX AAAY96528;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 9.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.

XX Synthetic.

XX Key Location/Qualifiers

XX Modified-site 1 /note= "optionally linked to an Fc molecule"

XX Peptide 6..19

XX Peptide 20..27

XX Peptide 28..41

XX /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX

PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

PS Claim 16; Page 65; 9lpp: English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-TMP-2],
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
 CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A, F, M, or K; X-1-1 = A, I, V,
 CC L, F, S, T, K, H, or E; X-1-2 = A, I, V, L, F, G, S, or Q; X-1-3 = R, K,
 CC T, V, N, Q or G; X-1-4 = A, I, V, L, F, T, R, E, or G; L-1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 41 AA;

Query Match 84.2%; Score 144; DB 21; Length 41;

Best Local Similarity 83.3%; Pred. No. 1.6e-13;

Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAARA---GPNIEGPTLRQWLAARA 32

DB 6 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLAARA 41

RESULT 21

AAAB17281

ID AAAB17281 standard; Peptide; 42 AA.

XX AAAB17281;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:337.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX

XX Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

XX autoimmune diseases -

XX

XX Disclosure; Page 313; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ Sequence 42 AA;

Query Match 84.2%; Score 144; DB 21; Length 42;

Best Local Similarity 83.3%; Pred. No. 1.7e-13;

Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIGEGPTLRQWLAAARA 32

|||||

Db 7 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

|||||

RESULT 22

AAB17282

ID AAB17282 standard; Peptide; 42 AA.

XX

AC AAB17282;

XX

DT 31-OCT-2000 (first entry)

XX

DE TPO-mimetic peptide sequence SEQ ID NO:338.

XX

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

XX

OS Synthetic.

OS

PN WO2000024782-A2.

XX

PD 04-MAY-2000.

XX

PF 25-OCT-1999; 99WO-US25044.

XX

XX 23-OCT-1998; 98US-0105371.

PR

PR 22-OCT-1999; 99US-0428082.

XX

PA (AMGE-) AMGEN INC.

XX

PI Feige U, Liu C, Cheetham J, Boone TC;

XX

DR WPI; 2000-350702/30.

XX

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX

PS Disclosure; Page 313; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ Sequence 42 AA;

Query Match 84.2%; Score 144; DB 21; Length 42;

Best Local Similarity 83.3%; Pred. No. 1.7e-13;

Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIGEGPTLRQWLAAARA 32

|||||

Db 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

|||||

RESULT 23

AAB17308

ID AAB17308 standard; Peptide; 42 AA.

XX

AC AAB17308;

XX

DT 31-OCT-2000 (first entry)

XX

DE Synthetic TWP-TMP gene construction peptide SEQ ID NO:374.

XX

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

XX

OS Homo sapiens.

OS Synthetic.

PN WO2000024782-A2.

XX

PD 04-MAY-2000.

XX

PF 25-OCT-1999; 99WO-US25044.

XX

XX 23-OCT-1998; 98US-0105371.

PR

PR 22-OCT-1999; 99US-0428082.

XX

PA (AMGE-) AMGEN INC.

XX

PI Feige U, Liu C, Cheetham J, Boone TC;

XX

DR WPI; 2000-350702/30.

XX

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX

PS Example 2; Page 327; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-Fl-(X2)b, where: Fl = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1-(L2)d-P2, (L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX SQ Sequence 42 AA;
 Query Match 84.2%; Score 144; DB 21; Length 42;
 Best Local Similarity 83.3%; Pred. No. 1.7e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAARA----GPNIGIEGPTLRQWLAARA 32
 |||||
 Db 7 IEGPTLRQWLAARAAGGGGGGIEGPTLRQWLAARA 42

RESULT 24

AAAY96530
 ID AAY96530 standard; Protein; 42 AA.

XX AC AAY96530;

XX DT 04-SEP-2000 (first entry)

XX DE Thrombopoietin mimetic peptide.

XX KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX OS Synthetic.

XX PN WO200024770-A2.

XX PD 04-MAY-2000.

XX PF 22-OCT-1999; 99WO-US24834.

XX PR 23-OCT-1998; 98US-0105348.

XX PA (AMGE-) AMGEN INC.

XX PI Liu C, Feige U, Cheetham J;

XX DR WPI: 2000-365108/31.

XX DR N-PSDB; AAA29225.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Example 2A; Page 48; 91pp; English.

XX Overlapping oligonucleotides were used to construct a synthetic
 CC gene encoding a thrombopoietin mimetic peptide (TMP), which
 CC was then fused in-frame to the Fc region of the human IgG1 chain (see
 CC AAY96529). A compound which binds to an mpl receptor comprising a TMP
 CC dimer joined by a linker [TMP_L1-(L1)-nTMP_2], is new. TMP_L1 and TMP_2
 CC are amino acid sequences varying from at least 10 to 14 residues in

CC length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2, X_2-X_1_3, X_2-X_1_4,
 CC X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and X_1-X_1_4. X_1 = I, A,
 CC V, L, S or R; X_2 = E, D, K or V; X_3 = G or A; X_4 = P; X_5 = T or S;
 CC X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N, or E; X_9 = W, Y or F;
 CC X_1_0 = L, I, V, A, E, M, or K; X_1_1 = A, I, V, L, F, S, T, K, H, or E;
 CC X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K, T, V, N, Q or G; X_1_4 =
 CC A, I, V, L, F, T, R, E, or G; L_1 = linker comprising 1 to 20 amino
 CC acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl
 CC receptor which mediates the activity of endogenous thrombopoietin. The
 CC TMPs are useful for increasing the production of platelets or platelet
 CC precursors (e.g. megakaryocytes) in a mammal, which is useful for
 CC treatment of diseases which involve thrombocytopenia, e.g. aplastic
 CC anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus
 CC associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 42 AA;

Query Match 84.2%; Score 144; DB 21; Length 42;
 Best Local Similarity 83.3%; Pred. No. 1.7e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAARA----GPNIGIEGPTLRQWLAARA 32

|||||
 Db 7 IEGPTLRQWLAARAAGGGGGGIEGPTLRQWLAARA 42

RESULT 25

AAB17311

ID AAB17311 standard; Peptide; 60 AA.

XX AC AAB17311;

XX DT 31-OCT-2000 (first entry)

XX DE Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI: 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases

XX Example 2; Page 331; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-Fl-(X2)b, where: Fl = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1-(L2)d-P2, (L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 60 AA;
 SQ
 Query Match 84.2%; Score 144; DB 21; Length 60;
 Best Local Similarity 83.3%; Pred. No. 2.5e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIGIEGPTLRQWLAAARA 32
 |||||
 DB 2 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 37

RESULT 26
 AAB16960
 ID AAB16960 standard; Protein; 269 AA.
 AC AAB16960;
 XX
 XX 31-OCT-2000 (first entry)
 DT
 XX
 DE TMP-TMP-Fc protein sequence SEQ ID NO:10.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Homo sapiens.
 OS Synthetic.

XX WO200024782-A2.
 XX 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 XX (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 DR WPI: 2000-350702/30.
 DR N-PSDB; AAA69446.
 XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 2; Page 185-186; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 269 AA;
 SQ
 Query Match 84.2%; Score 144; DB 21; Length 269;
 Best Local Similarity 83.3%; Pred. No. 1.4e-12;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARA----GPNIGIEGPTLRQWLAAARA 32
 |||||
 DB 2 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 37

RESULT 27
 AAY96531
 ID AAY96531 standard; Protein; 269 AA.
 XX
 AC AAY96531;
 XX
 XX 04-SEP-2000 (first entry)
 DT
 XX
 DE Human IgG1 Fc TMP fusion protein.
 XX
 KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.
 XX
 OS Homo sapiens.
 XX WO200024770-A2.

XX 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24834.
 XX
 PR 23-OCT-1998; 98US-0105348.
 XX
 XX (AMGE-) AMGEN INC.
 XX
 PI Liu C, Feige U, Cheetham J;
 XX
 DR WPI: 2000-365108/31.
 DR N-PSDB; AAA29229.
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 PS Example 2A; Page 49-50; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TWP-1-(L1)-TWP-2],
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
 CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N,
 CC or E; X-9 = W, Y or F; X-1-0 = L, I, V, A, F, M, or K; X-1-1 = A, I, V,
 CC L, F, S, T, K, H, or E; X-1-2 = A, I, V, L, F, G, S, or Q; X-1-3 = R, K,
 CC T, V, N, Q or G; X-1-4 = A, I, V, L, F, T, R, E, or G; L-1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 269 AA;
 SQ Query Match 84.2%; Score 144; DB 21; Length 269;
 Best Local Similarity 83.3%; Pred. No. 1.4e-12;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 IEPTLRQWLAAARA----GPNIGIEPTLRQWLAAARA 32
 |||||
 Db 234 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAARA 269

RESULT 28
 AAB17294
 ID AAB17294 standard; Peptide: 37 AA.

XX AAB17294;
 AC AAB17294;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:350.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 37 AA;
 SQ Query Match 83.9%; Score 143.5; DB 21; Length 37;
 Best Local Similarity 81.1%; Pred. No. 1.7e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

QY 1 IEPTLRQWLAAARA----GPNIGIEPTLRQWLAAARA 32
 |||||
 Db 1 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAARA 37

RESULT 29
 AAB17295
 ID AAB17295 standard; Peptide: 38 AA.

XX AAB17295;
 AC AAB17295;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:351.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 ; Search time 5,32084 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-23

Perfect score: 171

Sequence: 1 IEPTLRQWLAARAGPNGIEPTLRQWLAARA 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_patents_AA.*
2: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
7: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78.5	45.9	25	2	US-08-764-640-231
2	78.5	45.9	25	3	US-09-244-298A-231
3	78.5	45.9	25	4	US-09-516-704-231
4	73	42.7	14	2	US-08-764-640-13
5	73	42.7	14	2	US-08-764-640-193
6	73	42.7	14	3	US-08-973-225-13
7	73	42.7	14	3	US-08-973-225-193
8	73	42.7	14	3	US-09-244-298A-13
9	73	42.7	14	3	US-09-244-298A-193
10	73	42.7	14	4	US-09-516-704-13
11	73	42.7	14	4	US-09-516-704-193
12	73	42.7	15	2	US-08-764-640-17
13	73	42.7	15	2	US-08-764-640-185
14	73	42.7	15	3	US-08-973-225-17
15	73	42.7	15	3	US-08-973-225-185
16	73	42.7	15	3	US-09-244-298A-17
17	73	42.7	15	3	US-09-244-298A-185
18	73	42.7	15	4	US-09-516-704-17
19	73	42.7	15	4	US-09-516-704-185
20	73	42.7	16	2	US-08-764-640-18
21	73	42.7	16	2	US-08-764-640-194
22	73	42.7	16	2	US-08-764-640-232
23	73	42.7	16	3	US-08-973-225-18
24	73	42.7	16	3	US-08-973-225-194
25	73	42.7	16	3	US-08-973-225-220
26	73	42.7	16	3	US-09-244-298A-18
27	73	42.7	16	3	US-09-244-298A-194

28 73 42.7 16 3 US-09-244-298A-232
29 73 42.7 16 4 US-09-516-704-18
30 73 42.7 16 4 US-09-516-704-194
31 73 42.7 16 4 US-09-516-704-232
32 69 40.4 14 2 US-08-764-640-195
33 69 40.4 14 2 US-08-764-640-199
34 69 40.4 14 3 US-08-973-225-195
35 69 40.4 14 3 US-08-973-225-199
36 69 40.4 14 3 US-09-244-298A-195
37 69 40.4 14 3 US-09-244-298A-199
38 69 40.4 14 4 US-09-516-704-195
39 69 40.4 14 4 US-09-516-704-199
40 69 40.4 15 2 US-08-764-640-196
41 69 40.4 15 2 US-08-764-640-200
42 69 40.4 15 2 US-08-764-640-209
43 69 40.4 15 2 US-08-764-640-215
44 69 40.4 15 3 US-08-973-225-196
45 69 40.4 15 3 US-08-973-225-200

ALIGNMENTS

RESULT 1
US-08-764-640-231
; Sequence 231, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yio, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:

NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
US-08-764-640-231

Query Match 45.9%; Score 78.5; DB 2; Length 25;
Best Local Similarity 46.4%; Pred. No. 0.00027;
Matches 13; Conservative 8; Mismatches 2; Indels 5; Gaps 1;

QY 2 EGPTRLQWLAARAGPNGIEGPTLRQWLA 29
:|||||:|:| :|||||:|:|
DB 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 2

US-09-244-298A-231
Sequence 231, Application US/09244298A
Patent No. 6121238

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"

US-09-244-298A-231

Query Match 45.9%; Score 78.5; DB 3; Length 25;
Best Local Similarity 46.4%; Pred. No. 0.00027;
Matches 13; Conservative 8; Mismatches 2; Indels 5; Gaps 1;

QY 2 EGPTRLQWLAARAGPNGIEGPTLRQWLA 29

Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 3

US-09-516-704-231
Sequence 231, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 231:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
SEQUENCE DESCRIPTION: SEQ ID NO: 231:

US-09-516-704-231

Query Match 45.9%; Score 78.5; DB 4; Length 25;
Best Local Similarity 46.4%; Pred. No. 0.00027;
Matches 13; Conservative 8; Mismatches 2; Indels 5; Gaps 1;

QY 2 EGPTRLQWLAARAGPNGIEGPTLRQWLA 29

Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 4

US-08-764-640-13
Sequence 13, Application US/08764640
Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 42.7%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 5
US-08-764-640-193
Sequence 193, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193

Query Match 42.7%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAARA 14

RESULT 6
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherril S.
APPLICANT: Mattheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A

FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 42.7%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
|||||
DB 1 IEGPTLRQWLAAARA 14

RESULT 7
US-08-973-225-193
; Sequence 193, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Haselden, Sherril S.
; Matheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-08-973-225-193

Query Match 42.7%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
|||||
DB 1 IEGPTLRQWLAAARA 14

RESULT 8
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 11-Dec-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-13

Query Match 42.7%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
|||||
DB 1 IEGPTLRQWLAAARA 14

RESULT 9
US-09-244-298A-193
; Sequence 193, Application US/09244298A

Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-193
; Query Match 42.7%; Score 73; DB 3; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.00075;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; QY 1 IEPTLROWLAARA 14
; Db 1 IEPTLROWLAARA 14
; RESULT 10
; US-09-516-704-13
; Sequence 13, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR

NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-516-704-13
; Query Match 42.7%; Score 73; DB 4; Length 14;
; Best Local Similarity 100.0%; Pred. No. 0.00075;
; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; QY 1 IEPTLROWLAARA 14
; Db 1 IEPTLROWLAARA 14
; RESULT 11
; US-09-516-704-193
; Sequence 193, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-09-516-704-193
Query Match 42.7%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00075;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
|||||

Db 1 IEGPTLRQWLAARA 14

RESULT 12

US-08-764-640-17
Sequence 17, Application US/08764640

Patent No. 5869451

Patent No. 5869451 5837683

GENERAL INFORMATION:

APPLICANT: Dower, William J.

APPLICANT: Barrett, Ronald W.

APPLICANT: Cwiria, Steven E.

APPLICANT: Gates, Christian

APPLICANT: Schatz, Peter J.

APPLICANT: Balasubramanian, Palaniappan

APPLICANT: Wagstrom, Christopher R.

APPLICANT: Hendren, Richard W.

APPLICANT: Deprince, Randolph B.

APPLICANT: Podduturi, Surekha

APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-17

Query Match 42.7%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
|||||

Db 1 IEGPTLRQWLAARA 14

RESULT 13

US-08-764-640-185

Sequence 185, Application US/08764640

Patent No. 5869451

Patent No. 5869451 5837683

GENERAL INFORMATION:

APPLICANT: Dower, William J.

APPLICANT: Barrett, Ronald W.

APPLICANT: Cwiria, Steven E.

APPLICANT: Gates, Christian

APPLICANT: Schatz, Peter J.

APPLICANT: Balasubramanian, Palaniappan

APPLICANT: Wagstrom, Christopher R.

APPLICANT: Hendren, Richard W.

APPLICANT: Deprince, Randolph B.

APPLICANT: Podduturi, Surekha

APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-764-640-185

Query Match 42.7%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
|||||

Db 2 IEGPTLRQWLAARA 15

RESULT 14

US-08-973-225-17
; Sequence 17, Application US/08973225A
; Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:

US-08-973-225-17

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLROWLAARA 14

Db 1 IEPTLROWLAARA 14

RESULT 15

US-08-973-225-185
; Sequence 185, Application US/08973225A
; Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.

Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-08-973-225-185

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLROWLAARA 14

Db 2 IEPTLROWLAARA 15

RESULT 16

US-09-244-298A-17
; Sequence 17, Application US/09244298A
; Patent No. 6121238

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprience, Randolph B.
Podduturi, Surekha
Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14
|||||

RESULT 17
US-09-244-298A-185
Sequence 185, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-185

Query Match 42.7%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15
|||||

RESULT 18
US-09-516-704-17
Sequence 17, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match 42.7%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 1 IEGPTLRQWLAARA 14
|||||

Db 1 IEGPTLRQWLAARA 14

RESULT 19

US-09-516-704-185

; Sequence 185, Application US/09516704

; Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwirla, Steven E.

Gates, Christian

Schatz, Peter J.

Balasubramanian, Palaniappan

Wagstrom, Christopher R.

Hendren, Richard W.

Deprince, Randolph B.

Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516.704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-09-516-704-185

Query Match 42.7%; Score 73; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.00081;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14

Db 2 IEGPTLRQWLAARA 15

RESULT 20

US-08-764-640-18

; Sequence 18, Application US/08764640

; Patent No. 5869451

; Patent No. 5869451 5837683

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwirla, Steven E.

Gates, Christian

Schatz, Peter J.

Balasubramanian, Palaniappan

Wagstrom, Christopher R.

Hendren, Richard W.

Deprince, Randolph B.

Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product= "Beta-ala"

US-08-764-640-18

Query Match 42.7%; Score 73; DB 2; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.00087;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14

Db 1 IEGPTLRQWLAARA 14

RESULT 21

US-08-764-640-194

; Sequence 194, Application US/08764640

; Patent No. 5869451

; Patent No. 5869451 5837683

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwirla, Steven E.

Gates, Christian

Schatz, Peter J.

Balasubramanian, Palaniappan

Wagstrom, Christopher R.

Hendren, Richard W.

Deprince, Randolph B.

Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product= "Beta-ala"

US-08-764-640-18

Query Match 42.7%; Score 73; DB 2; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.00087;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14

Db 1 IEGPTLRQWLAARA 14

Wed Oct 9 10:29:33 2002

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: peptide
US-08-764-640-194

Query Match 42.7%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLQWLAA 14
DB 2 IEPTLQWLAA 15

RESULT 22
US-08-764-640-232
; Sequence 232, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depndence, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: peptide
US-08-764-640-232

Query Match 42.7%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLQWLAA 14
DB 2 IEPTLQWLAA 15

RESULT 23
US-08-973-225-18
; Sequence 18, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherril S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide

```



```
;
;
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Beta-ala"
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18

Query Match          42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 1 IEGPTLQWLAAARA 14

RESULT 24
US-08-973-225-194
; Sequence 194, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirila, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Matheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
;
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
;
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194

Query Match          42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 2 IEGPTLQWLAAARA 15
```

```
RESULT 25
US-08-973-225-220
; Sequence 220, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirila, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Matheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
;
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; THROMBOPOIETIN RECEPTOR
;
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 220:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220

Query Match          42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLQWLAAARA 14
   |||||
DB 2 IEGPTLQWLAAARA 15

RESULT 26
US-09-244-298A-18
; Sequence 18, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirila, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
```

APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
US-09-244-298A-18

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 1 IEGPTLRQWLAAARA 14
|||||
Db 1 IEGPTLRQWLAAARA 14

RESULT 27
US-09-244-298A-194
Sequence 194, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depvince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 1 IEGPTLRQWLAAARA 14
|||||
Db 2 IEGPTLRQWLAAARA 15

RESULT 28
US-09-244-298A-232
Sequence 232, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depvince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-232

Query Match 42.7%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTTLQWLAAARA 14
DB 2 IEPTTLQWLAAARA 15

RESULT 29

US-09-516-704-18
Sequence 18, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product="Beta-ala"

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-516-704-18

Query Match 42.7%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTTLQWLAAARA 14

DB 1 IEPTTLQWLAAARA 14

RESULT 30

US-09-516-704-194
Sequence 194, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 194:

US-09-516-704-194

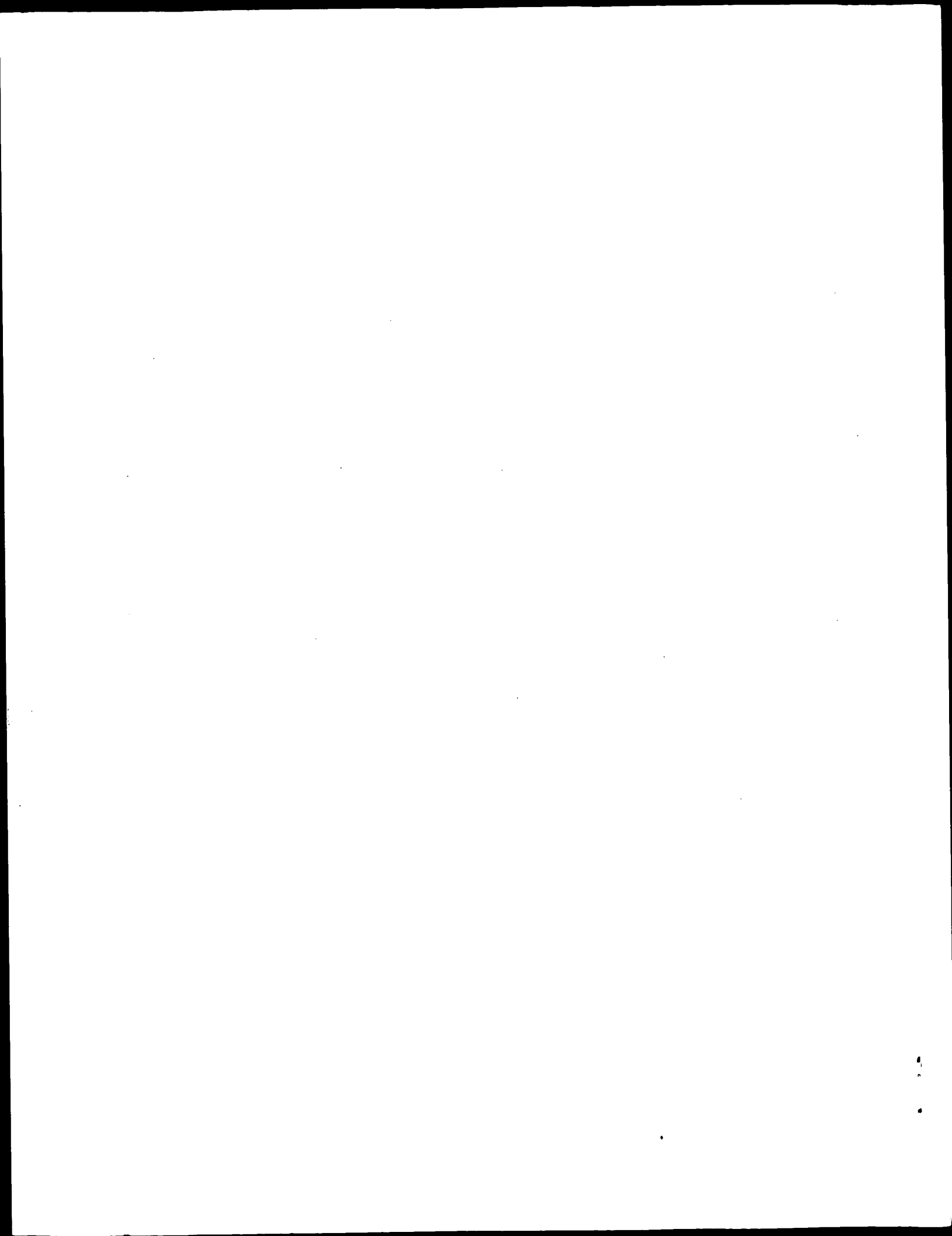
Query Match 42.7%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.00087;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTTLQWLAAARA 14

DB 2 IEPTTLQWLAAARA 15

Search completed: October 9, 2002, 09:06:29

Job time : 5.32084 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 7.19438 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838C-23
 Perfect score: 171
 Sequence: 1 IEPTLROWLAARACNGIEPTLROWLAAR 32

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database :      PIR_71:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	63	36.8		683	2	B71325	conserved hypothet
2	56	32.7		361	2	F87286	cation efflux fami
3	55	32.2		420	2	JW0076	sorbitol oxidase (
4	54	31.6		346	2	D85818	unknown protein en
5	54	31.6		1095	2	B83471	probable pyruvate
6	53	31.0		296	2	AG0147	probable membrane
7	53	31.0		326	2	C75350	probable UV damage
8	53	31.0		524	1	VGNCV	spike glycoprotein
9	52.5	30.7		814	2	G02390	disintegrin-like m
10	52	30.4		396	2	C35254	conserved hypothet
11	52	30.4		440	2	S65358	familial Alzheimer
12	51.5	30.1		371	2	F83487	hypothetical prote
13	51	29.8		281	2	G72680	hypothetical prote
14	51	29.8		306	2	D70601	UTP-glucose-1-pho
15	51	29.8		589	2	T29299	hypothetical prote
16	51	29.8		600	2	C83221	transport protein
17	51	29.8		697	1	S04987	STIS-binding prote
18	51	29.8		719	2	B95325	conserved hypothet
19	50	29.2		150	2	A83634	nitric-oxide reduc
20	50	29.2		351	2	C75479	conserved hypothet
21	50	29.2		410	1	DEPSX	3-methyl-2-oxobuta
22	50	29.2		410	2	C83365	2-oxoisovalerate d
23	50	29.2		415	2	T38324	probable trna meth
24	50	29.2		460	2	S06469	photosystem II chl
25	50	29.2		472	2	T20454	hypothetical prote
26	50	29.2		904	2	C70559	probable poia prot
27	49.5	28.9		333	2	A36925	transcription acti
28	49.5	28.9		341	2	A13083	monooxygenase (imp
29	49.5	28.9		355	2	H98202	hypothetical prote

30	49	28.7	214	2	T22896	hypothetical prote
31	49	28.7	369	1	DEBSPF	pyruvate dehydroge
32	49	28.7	385	2	A44102	di-N-acetylchitobi
33	49	28.7	403	2	AD0748	tyrosine-specific
34	49	28.7	425	2	A83032	probable permease
35	49	28.7	459	2	S42647	photosystem II chl
36	49	28.7	459	2	AD2342	photosystem II CP4
37	49	28.7	735	2	T50630	alpha subunit of c
38	49	28.7	992	2	A83324	probable sensor/re
39	49	28.7	1024	2	S18251	collagen alpha 1(X
40	49	28.7	3198	2	A43426	collagen alpha 2(X
41	48.5	28.4	209	2	C87617	glutathione S-tran
42	48.5	28.4	493	2	T48219	hypothetical prote
43	48	28.1	72	1	RSBPXL	excisionase - phag
44	48	28.1	72	2	S06533	excisionase - phag
45	48	28.1	72	2	A90729	excisionase limpor

ALIGNMENTS

RESULT 1
B71325
conserved hypothetical protein TP0421 - syphilis spirochete
C.Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C.Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C.Accession: B71325
C.R.Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G.
L.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
th, L.; Smith, H.O.; Venter, J.C.

```
Query Match      36.8%; Score 63; DB 2; Length 683;
Best Local Similarity 46.4%;
Matches 13; Conservative 3; Mismatches 12; Indels 0;
Gaps 0;
```

QY 4 PTLRQLAARAGPNIGIEPTLRQLAAR 31
| : : | : | | | | |
Db 74 PIILFIWIGNAVVRSGTIEGAALHOWGAAR 101

RESULT 2
F87286
cation efflux family protein [imported] - Caulobacter crescentus
C.Species: Caulobacter crescentus
C.Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C.Accession: F87286
R.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko-
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A.Title: Complete Genome Sequence of Caulobacter crescentus.
A.Reference number: A87249; MUID:21173698; PMID:11259647

Query Match	32.7%	Score 56;	DB 2;	Length 361;
Best Local Similarity	54.5%	Pred. No. 8.2;		

Matches 15; Conservative 4; Mismatches 6; Indels 12; Gaps 2;

QY 2 EGPTRLROWLAARAGPNCIEGPTLRQ 26
 Db 249 EDPSVREWLRLARATWQPEWQVHLSNGIEGPQDRR 285

RESULT 8
 VGNCV
 conserved hypothetical protein G precursor - Chandipura virus
 C:Species: Chandipura virus
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
 C:Accession: A32443
 R:Masters, P.S.; Bella, R.S.; Butcher, M.; Patel, B.; Ghosh, H.P.; Banerjee, A.K.
 Virology 171, 285-290, 1989
 A:Title: Structure and expression of the glycoprotein gene of Chandipura virus.
 A:Reference number: A32443; MUID:89299473
 A:Accession: A32443
 A:Molecule type: mRNA
 A:Residues: 1-524 <MAS>
 A:Cross-references: GB:J04350; NID:G323376; PIDN:AAA42916.1; PID:G323377
 C:Genetics:

A:Gene: G
 A:Superfamily: rhabdovirus spike glycoprotein G
 C:Keywords: glycoprotein; spike protein; transmembrane protein
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-524/Product: spike glycoprotein G #status predicted <SGG>
 F:472-491/Domain: transmembrane #status predicted <TMN>
 F:184,344/Binding site: carbonylate (Asn) (covalent) #status predicted

Query Match 31.0%; Score 53; DB 1; Length 524;
 Best Local Similarity 37.0%; Pred. No. 29;
 Matches 10; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 1 IEGPTLROWLAARAGPNCIEGPTLRQ 27
 Db 359 IDGPVLKEPKGKRESGSISSDIWTQW 385

RESULT 9
 G02390
 disintegrin-like metalloproteinase MDC15 (EC 3.4.24.-) - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 31-Dec-2000
 C:Accession: G02390; PC4263
 R:Herren, B.; Raines, E.W.; Ross, R.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: H01157
 A:Accession: G02390
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-814 <HER>
 A:Cross-references: EMBL:U46005; NID:G1335871; PIDN:AAC51112.1; PID:G1335872
 R:McKie, N.; Edwards, T.; Dallas, D.J.; Houghton, A.; Stringer, B.; Graham, R.; Russell,
 Biochem. Biophys. Res. Commun. 230, 335-339, 1997
 A:Title: Expression of members of a novel membrane linked metalloproteinase family (ADAM
 A:Reference number: PC4263; MUID:97168971

A:Accession: PC4263
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-461 <MCK>
 A:Experimental source: articular chondrocyte
 C:Comment: This protein is a membrane bound protein and involved in cell/cell and cell/m
 C:Superfamily: mouse meltrin alpha; disintegrin homology
 C:Keywords: hydrolase; metalloproteinase; zinc
 F:420-503/Domain: disintegrin homology <DIS>
 F:348,352,358/Binding site: zinc (His) #status predicted
 F:349/Active site: Glu #status predicted

Query Match 30.7%; Score 52.5; DB 2; Length 814;
 Best Local Similarity 44.8%; Pred. No. 53;
 Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY 1 IEGPTLROWLAARAGPNCIEGPTLRQ 29
 Db 728 LKGPCTC-QYRAAQSGSERPGPPQALLA 755

RESULT 10
 T35254
 conserved hypothetical protein SC5F2A.12c - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 18-Aug-2000
 C:Accession: T35254
 R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, April 1999
 A:Reference number: Z21573
 A:Accession: T35254
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-396 <OLI>
 A:Cross-references: EMBL:AL049587; PIDN:CAB40679.1; GSPDB:GN00070; SCOEEDB:SC5F2A.12c
 A:Experimental source: strain A3(2)
 C:Genetics:

A:Gene: SCOEEDB:SC5F2A.12c
 C:Superfamily: Streptomyces coelicolor conserved hypothetical protein SC5F2A.12c

Query Match 30.4%; Score 52; DB 2; Length 396;
 Best Local Similarity 34.4%; Pred. No. 29;
 Matches 11; Conservative 5; Mismatches 10; Indels 6; Gaps 2;

QY 4 PTLQWLAAARAGPNCIEGPTLRQ 31
 Db 293 PPARRWLSGLAPG--EGPSAERRAKSWFSVR 322

RESULT 11
 S65358
 familial Alzheimer's disease protein 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 25-Apr-1997
 C:Accession: S65358
 R:Matsumoto, A.; Matsumoto, R.; Fujiwara, Y.
 Eur. J. Biochem. 230, 337-343, 1995
 A:Title: Molecular cloning of human cDNA with a sequence highly similar to that of th
 A:Reference number: S65358; MUID:95324544
 A:Accession: S65358
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-440 <MAT>

Query Match 30.4%; Score 52; DB 2; Length 440;
 Best Local Similarity 46.7%; Pred. No. 33;
 Matches 14; Conservative 1; Mismatches 9; Indels 6; Gaps 1;

QY 3 GPTLROWLAARAGPNCIEGPTLRQ 32
 Db 374 GPDLSALAGRVGTGTF-----PFSAA 397

RESULT 12
 F83487
 hypothetical protein PA1267 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83487
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337
 A:Accession: F83487
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-371 <STO>

```

Query Match          29.8%; Score 51; DB 2; Length 306;
Best Local Similarity 69.2%; Pred. No. 30;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
. 3 GPTLRQWLAARAG 15

```

```

Query Match      29.8
Best Local Similarity 52.9
Matches          9; Conservative

```

RESULT 17
4987
TS-binding protein spi05 - F
Species: Torpedo californica
Date: 30-Jun-1993 #sequence
Accession: S04987; S30070
Jentsch, T.J.; Garcia, A.M.;
Biochem. J. 261, 155-166, 1989

hypothetical protein AGR_L1143 [imported] - Agrobacterium tumefaciens (strain C58,
C; Species: Agrobacterium tumefaciens
C; Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C; Accession: H98202

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*
 A:Reference number: A97359; PMID:11743194
 A:Accession: H98202
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-355 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK89146.1; PID:g15158956; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L1143
 A:Map position: linear chromosome

Query Match 28.9%; Score 49.5; DB 2; Length 355;
 Best Local Similarity 35.1%; Pred. No. 55;
 Matches 13; Conservative 4; Mismatches 13; Indels 7; Gaps 2;

QY 2 EGPTRLQWLAAAR-----AGPNGIE--GPTLRQWLAAAR 31
 DB 226 QGQSPWIAANMEGRFVYPNGLERLAAQARDWTAAR 262

RESULT 30

T22896

hypothetical protein F58B3.3 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T22896

R:Harris, B.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z19633

A:Accession: T22896

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-214 <WIL>

A:Cross-references: EMBL:Z73427; PIDN:CAA97801.1; GSPDB:GN00022; CESP:F58B3.3

A:Experimental source: Clone F58B3

C:Genetics:

A:Gene: CESP:F58B3.3

A:Map position: 4

A:Introns: 68/1

C:Superfamily: *Caenorhabditis elegans* hypothetical protein F58B3.3

Query Match 28.7%; Score 49; DB 2; Length 214;
 Best Local Similarity 50.0%; Pred. No. 38;
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 PTLRQWLAAARAGPNGI 19

DB 191 PTHQWEGTAGPCGV 206

Search completed: October 9, 2002, 09:05:00
 Job time : 7.19438 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 3.82201 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-23

Perfect score: 1/1

Sequence: 1 IEPTLRQWLARAGPNIGETPLRQWLARA 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.5	31.3	266	1 SC02_HUMAN	O43819 homo sapien
2	53	31.0	524	1 VGIG_CHAV	P13180 chandipura
3	52.5	30.7	814	1 AD15_HUMAN	O13444 homo sapien
4	51	29.8	696	1 SP15_TORCA	P19965 torpeda cal
5	50	29.2	243	1 DREFM_HUMAN	Q9hbi1 homo sapien
6	50	29.2	246	1 TONB_PASHA	P72204 pasteurella
7	50	29.2	410	1 ODBA_PSEPU	P09060 pseudomonas
8	50	29.2	415	1 TRMO_SCHPO	O13947 schizosacch
9	50	29.2	904	1 DPOL_MYCTU	P09193 synecocyst
10	50	29.2	904	1 PSBC_SCHPO	O07700 mycobacteri
11	49.5	28.9	333	1 CBBR_XANFL	P25545 xanthobacte
12	49	28.7	368	1 ODBA_BACST	P21873 bacillus st
13	49	28.7	385	1 DIAC_HUMAN	O01459 homo sapien
14	49	28.7	735	1 CNGL_CHICK	O28083 gallus gall
15	49	28.7	911	1 CALB_BOVIN	P81615 mus musculu
16	48.5	28.4	122	1 UROC_MOUSE	P11883 bacterioph
17	48	28.1	72	1 VXIS_BP434	P03699 bacterioph
18	48	28.1	72	1 VXIS_LAMBD	O9kq28 vibrio chol
19	48	28.1	270	1 VL76_VIBCH	O9cbu0 mycobacteri
20	48	28.1	297	1 XERC_MYCLE	P21881 bacillus su
21	48	28.1	370	1 ODBA_BACSU	P08123 homo sapien
22	48	28.1	1366	1 CA21_HUMAN	O9yft9 aeropyrum p
23	47.5	27.8	562	1 SYK_AERPE	O75474 homo sapien
24	47	27.5	113	1 FRT2_HUMAN	O06236 mycobacteri
25	47	27.5	357	1 PYR2_MYCTU	P14643 pinus thunb
26	47	27.5	473	1 PSBC_PINTH	O15067 homo sapien
27	47	27.5	1338	1 PUR4_HUMAN	O01149 mus musculu
28	47	27.5	1372	1 FUR4_MOUSE	P33479 pseudorabie
29	47	27.5	1446	1 IE18_MOUSE	P11675 pseudorabie
30	47	27.5	1461	1 IE18_PRVKA	O64612 rattus norv
31	47	27.5	1711	1 PTPO_RAT	P12107 homo sapien
32	47	27.5	1806	1 CALB_HUMAN	O10815 mycobacteri
33	46	26.9	298	1 XERC_MYCTU	

34	46	26.9	335	1	FABH_MYCTU	O06399 mycobacteri
35	46	26.9	369	1	CAL2_CHICK	P02460 gallus gall
36	46	26.9	503	1	PSBB_ODOSI	P49471 odontella s
37	46	26.9	568	1	G6P1_ODOSI	P54234 clarkia arc
38	46	26.9	568	1	G6P1_CLAAR	P54236 clarkia fra
39	46	26.9	568	1	G6P1_CLAWI	P54240 clarkia wil
40	46	26.9	568	1	G6P1_CLAXA	P54240 clarkia xan
41	46	26.9	568	1	G6P1_OENME	P54243 oenothera m
42	46	26.9	569	1	G6P1_CLACO	P54235 clarkia con
43	46	26.9	569	1	G6P1_CLALE	P34796 clarkia lew
44	46	26.9	570	1	G6P1_CLARO	P54238 clarkia ros
45	46	26.9	589	1	PHBC_ALCEU	P23608 a poly-beta

ALIGNMENTS

RESULT 1
SC02_HUMAN STANDARD; PRT; 266 AA.
AC O43819; Q9UR87;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SC02 protein homolog, mitochondrial precursor.
GN SC02.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Monocytes;
RA Smink L.J., Burton J.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS FIC LYS-140 AND PHE-225.
RX MEDLINE=20014747; PubMed=10545952;
RA Papadopoulos L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,
RA Sadlock J.E., Krishna S., Walker W., Selby J., Glerum D.M.,
RA Van Coster R., Lyon G., Scalais E., Lebel R., Kaplan P., Shanske S.,
RA De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.;
RT "Fatal infantile cardioencephalomyopathy with COX deficiency and
mutations in SC02, a COX assembly gene.";
RL Nat. Genet. 23:333-337(1999).
CC -!- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER
TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.
CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- DISEASE: DEFECTS IN SC02 ARE THE CAUSE OF FATAL INFANTILE
CARDIOENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS
CHARACTERIZED BY HYPERTROPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND
GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX
ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX
DEFICIENCIES.
CC -!- SIMILARITY: BELONGS TO THE SC01/2 FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AF177385; AA05313.1; -
EMBL: AL021683; CAA16671.1; -
MIM: 604272; -
MIM: 604377; -
MIM: 220110; -
InterPro: IPR003782; SC01_Senc.
Pfam: PF02630; SC01-Senc; 1.
Mitochondrion; Transit peptide; Disease mutation; Polymorphism.

FT	CARBOHYD	344	344	N-LINKED (GLCNAC...)	(POTENTIAL).
FT	LIPID	479	479	PALMITATE (POTENTIAL).	
FT	LIPID	484	484	PALMITATE (POTENTIAL).	
FT	SEQUENCE	524 AA;	58826 MW;	A84AF0A5FFFB73CD CRC64;	
Query Match		31.0%;	Score 53;	DB 1;	Length 524;
Best Local Similarity		37.0%;	Pred. No. 11;		
Matches		10;	Conservative	4;	Mismatches 13; Indels 0; Gaps 0;
QY	1	IEGPTLRWLAAARAGPNGIEGPTLRQW	27.		
Db	359	IDGVLKEPKGRSPSGISSDIWTQW	385		
RESULT 3					
AD15_HUMAN		STANDARD;	PRT;	814 AA.	
ID	AD15_HUMAN				
AC	Q13444;	Q13493;			
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	01-MAR-2002	(Rel. 41, Last annotation update)			
DE	ADAM 15 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase domain 15) (Metalloproteinase-like, disintegrin-like, and cysteine-rich protein 15) (MDC-15) (Metalloprotease RGD disintegrin protein) (Metargidin).				
DE	ADAM15 OR MDC15.				
GN	Homo sapiens (Human).				
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OC	NCBI_TaxID=9606;				
OX	[1]				
RN	SEQUENCE FROM N.A.				
RP	TISSUE=Breast Carcinoma;				
RC	MEDLINE=96214870; PubMed=8617717;				
RX	Kraetzschmar J., Lum L., Blobel C.P.;				
RA	"Metargidin, a membrane-anchored metalloprotease-disintegrin protein with an RGD integrin binding sequence.";				
RT	J. Biol. Chem. 271:4593-4596(1996).				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RP	TISSUE=Umbilical vein;				
RC	MEDLINE=97192141; PubMed=9039960;				
RX	Herten B., Raines E.W., Ross R.;				
RA	"Expression of a disintegrin-like protein in cultured human vascular cells and in vivo.";				
RT	FASEB J. 11:173-180(1997).				
RL	[3]				
RN	INTERACTION WITH INTEGRIN ALPHA-V-BETA3.				
RP	MEDLINE=98184837; PubMed=9516430;				
RX	Zhang X.P., Kamata T., Yokoyama K., Puzon-McLaughlin W., Takada Y.;				
RA	"Specific interaction of the recombinant disintegrin-like domain of MDC-15 (metargidin, ADAM-15) with integrin alphavbeta3.";				
RT	J. Biol. Chem. 273:7345-7350(1998).				
RL	[-]				
CC	FUNCTION: MAY BE INVOLVED IN CELL-SURFACE PROTEOLYSIS, CELL ADHESION OR INTRACELLULAR PROTEIN MATURATION.				
CC	[-] COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).				
CC	[-] SUBUNIT: INTERACTS WITH INTEGRIN ALPHA-V-BETA3, ENDOPHILIN 1 AND SORTING NEXIN 9. ENDOPHILIN 1 AND SORTING NEXIN 9 PREFERENTIALLY BIND THE PRECURSOR BUT NOT THE PROCESSED FORM OF ADAM15, SUGGESTING THAT THE INTERACTION OCCURS IN A SECRETORY PATHWAY COMPARTMENT PRIOR TO THE MEDIAL GOLGI (BY SIMILARITY).				
CC	[-] SUBCELLULAR LOCATION: Type 1 membrane protein.				
CC	[-] TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED. OVEREXPRESSED IN ARTEROSCLEROTIC LESIONS. CONSTITUTIVELY EXPRESSED IN CULTURED ENDOTHELIAL AND SMOOTH MUSCLE.				
CC	[-] DOMAIN: THE CYTOPLASMIC DOMAIN INTERACTS WITH ENDOPHILIN 1 AND SORTING NEXIN 9 (BY SIMILARITY).				
CC	[-] DOMAIN: DESINTEGRIN DOMAIN BINDS TO INTEGRIN ALPHA-V-BETA3.				
CC	[-] PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).				
CC	[-] SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.				
CC	[-] SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.				
CC	[-] SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.				

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; U46005; AAC51112.1; -;
EMBL; U41767; AAC50404.1; -;
HSSP: PL8619; LFVL.
MEROPS: M12.215; -.
MIM: 605548; -.
InterPro: IPR001762; Disintegrin.
InterPro: IPR000561; EGF-like.
InterPro: IPR001818; Matrixin.
InterPro: IPR002870; Pep_M12B_propep.
InterPro: IPR001590; Repolysin.
InterPro: IPR000130; Zn_Mrpeptidse.
Pfam: PF00200; disintegrin; 1.
Pfam: PF01562; Pep_M12B_propep; 1.
Pfam: PF01421; Repolysin; 1.
ProDom: PD000664; Disintegrin; 1.
SMART: SM00050; DISIN; 1.
SMART: SM00181; EGF; 1.
PROSITE: PS00215; ADAM_MEPRO; 1.
PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
PROSITE: PS00214; DISINTEGRIN_2; 1.
PROSITE: PS00022; EGF_1; FALSE_NEG.
PROSITE: PS01186; EGF_2; 1.
PROSITE: PS00142; ZINC_PROTEASE; 1.
PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Transmembrane; EGF-like domain; SH3-binding.
SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 206 BY SIMILARITY.
FT CHAIN 207 814 ADAM 15.
FT DOMAIN 207 696 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 697 717 POTENTIAL.
FT DOMAIN 718 814 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 207 414 METALLOPROTEASE.
FT DOMAIN 421 508 DISINTEGRIN-LIKE.
FT DOMAIN 509 656 CYS-RICH.
FT DOMAIN 657 685 EGF-LIKE.
FT SITE 766 772 SH3-BINDING (POTENTIAL).
FT SITE 801 807 SH3-BINDING (POTENTIAL).
FT SITE 179 179 CYSTEINE SWITCH.
FT SITE 484 486 CELL ATTACHMENT SITE (POTENTIAL).
FT METAL 348 348 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 349 349 BY SIMILARITY.
FT METAL 352 352 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 358 358 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 323 409 POTENTIAL.
FT DISULFID 480 493 BY SIMILARITY.
FT DISULFID 657 667 BY SIMILARITY.
FT DISULFID 661 673 BY SIMILARITY.
FT DISULFID 675 684 BY SIMILARITY.
FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 606 606 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 714 714 G -> S (IN REF. 2).
FT CONFLICT 791 791 A -> P (IN REF. 2).
SEQUENCE 814 AA; 87686 MW; DD2EC26CB1314576 CRC64;
Query Match 30.7%; Score 52.5; DB 1; Length 814;
Best Local Similarity 44.8%; Pred No. 20;
Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;
QY 1 IEGPTLROWLAARAGPNIGEGPTLROWLA 29
::: 1: 11111: 11 1 11

AC Q9BHH1; 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Peptide deformylase, mitochondrial precursor (EC 3.5.1.88) (PDF)
 DE (Polypeptide deformylase).
 GN PDFIA OR PDF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20514156; PubMed=11060042;
 RA Giglion C., Serero A., Pierre M., Boisson B., Meinzel T.;
 RT Identification of eukaryotic peptide deformylases reveals
 RT universality of N-terminal protein processing mechanisms.";
 RL EMBO J. 19:5916-5929(2000).
 CC -|- FUNCTION: Removes the formyl group from the N-terminal Met of
 CC newly synthesized proteins (By similarity).
 CC -|- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
 CC methionyl peptide.
 CC -|- COFACTOR: Binds 1 iron(II) ion (By similarity).
 CC -|- SUBCELLULAR LOCATION: Mitochondrial (Potential).
 CC -|- TISSUE SPECIFICITY: Ubiquitous.
 CC -|- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF239156; AAC33968.1; -;
 CC EMBL; AF322879; AAK15624.1; -;
 CC InterPro: IPR000181; Pep_deformylase.
 CC Pfam: PF01327; Pep_deformylase; 1.
 CC ProDom: PD003844; Pep_deformylase; 1.
 CC protein biosynthesis; Hydrolase; Iron; Mitochondrion; Transit peptide.
 FT TRANSIT 1 2 MITOCHONDRION (POTENTIAL).
 FT CHAIN ? 243 PEPTIDE DEFORMYLASE.
 FT METAL 172 172 IRON (BY SIMILARITY).
 FT METAL 214 214 IRON (BY SIMILARITY).
 FT ACT_SITE 215 215 BY SIMILARITY.
 FT METAL 218 218 IRON (BY SIMILARITY).
 SQ SEQUENCE 243 AA; 27013 MW; B15A3456F0F8D689 CRC64;
 Query Match 29.2%; Score 50; DB 1; Length 243;
 Best Local Similarity 50.0%; Pred. No. 12; Indels 3; Gaps 0;
 Matches 8; Conservative 5; Mismatches 5;
 QY 11 AARAGPNCIEGPTLRO 26
 Db 31 SSTAAPDGVGEPALR 46
 RESULT 6
 TONB_PASHA
 ID TONB_PASHA STANDARD; PRT; 246 AA.
 AC P72204;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE TonB protein.
 GN TONB.
 OS Pasteurella haemolytica.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Mannheimia.
 OX NCBI_TaxID=75985;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROTYPE A1 / ATCC 43270;
 RA Graham M.R., Lo R.Y.C.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -|- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
 CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
 CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
 CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
 CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
 CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
 CC MEMBRANE PROTEINS (BY SIMILARITY).
 CC -|- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
 CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
 CC PERIPLASM (BY SIMILARITY).
 CC -|- SIMILARITY: BELONGS TO THE TONB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U62565; AAB09530.1; -;
 CC Transport; Protein transport; Inner membrane; Periplasmic;
 KW Transmembrane; Signal-anchor.
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 8 28 SIGNAL-ANCHOR (POTENTIAL).
 FT DOMAIN 29 246 PERIPLASMIC (POTENTIAL).
 SQ SEQUENCE 246 AA; 27785 MW; C9582F619FCA5B5 CRC64;
 Query Match 29.2%; Score 50; DB 1; Length 246;
 Best Local Similarity 47.4%; Pred. No. 13;
 Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
 QY 3 GPTLROWLAARAGPNCIEG 21
 Db 157 GPEIKQIGVAKAIPNAEG 175
 RESULT 7
 ODBA_PSEPU STANDARD; PRT; 410 AA.
 ID ODBA_PSEPU
 AC P09060;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 2-oxoisovalerate dehydrogenase alpha subunit (EC 1.2.4.4) (Branched-
 DE chain alpha-keto acid dehydrogenase component alpha chain (E1))
 DE (BKDH E1-alpha).
 DE GN BKDAL.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PPG2;
 RX MEDLINE=88329084; PubMed=3416875;
 RA Burns G., Brown T., Hatter K., Idriss J., Sokatch J.R.;
 RT "Similarity of the E1 subunits of branched-chain-oxoacid dehydrogenase
 RT from Pseudomonas putida to the corresponding subunits of mammalian
 RT branched-chain-oxoacid and pyruvate dehydrogenases.";
 RL Eur. J. Biochem. 176:311-317(1988).
 RN [2]
 RP SEQUENCE OF 1-17 FROM N.A.
 RC STRAIN=PPG2;
 RX MEDLINE=91008935; PubMed=2211503;


```

CC -----
DR EMBL; M21538; AA85378.1; -.
DR EMBL; D90909; BAA17799.1; -.
DR EMBL; X07018; CAA30071.1; -.
DR PIR; S06469; S06469.
DR PIR; S02380; S02380.
DR InterPro; IPR000932; PSII.
DR Pfam; PF00421; PSII; 1.
KW Photosynthesis; Photosystem II; Thylakoid; Chlorophyll;
KW Transmembrane; Complete proteome.
FT TRANSMEM 56 75 POTENTIAL.
FT TRANSMEM 107 129 POTENTIAL.
FT TRANSMEM 160 182 POTENTIAL.
FT TRANSMEM 202 224 POTENTIAL.
FT TRANSMEM 237 259 POTENTIAL.
FT TRANSMEM 269 291 POTENTIAL.
FT TRANSMEM 423 445 POTENTIAL.
FT TRANSMEM 54 55 R -> A (IN REF. 2).
FT CONFLICT 67 67 T -> N (IN REF. 3).
FT CONFLICT 162 162 Y -> I (IN REF. 3).
SQ SEQUENCE 472 AA; 51760 MW; D94D9FE7F3F66192D CRC64;

Query Match 29.2%; Score 50; DB 1; Length 472;
Best Local Similarity 35.0%; Pred. No. 24;
Matches 14; Conservative 3; Mismatches 13; Indels 10; Gaps 2;

QY 3 GPTLR-----QWLAARAGPNGIEGPTLRQ-----WLAARA 32
DB 352 GETMRFWDGPGWLEPLRGNGLDLKLRLQDPQWQVRRRA 391

RESULT 10
DPOL_MYCTU
ID POL_MYCTU STANDARD; PRT; 904 AA.
AC Q07700;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
GN POLA OR Rv1629 OR MT1665 OR MTC01B2.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=94124016; PubMed=8294019;
RA Mizrahi V., Huberts P., Dawes S.S., Dudding L.R.;
RT "A PCR method for the sequence analysis of the gyrA, polA and rnhA
RL gene segments from mycobacteria.";
RN [2]
RP SEQUENCE 136:287-250(1993).
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier C., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basmah D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moulie S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

```

```

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -1- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11920; AAB46393.1; -.
DR EMBL; Z95554; CAB08882.1; -.
DR EMBL; AE007030; AAK45935.1; -.
DR HSP; P19821; IBGX.
DR TIGR; MT1665; -.
DR TubercuList; Rv1629; -.
DR InterPro; IPR002562; 3_5_exonuclease.
DR InterPro; IPR002421; 5_3_exonuclease.
DR InterPro; IPR002298; DNA_polI.
DR InterPro; IPR001098; DNA_pol_A.
DR InterPro; IPR000513; EXO_N_I.
DR InterPro; IPR003583; HHH_1.
DR InterPro; IPR003584; HHH_2.
DR Pfam; PF01367; 5_3_exonuclease; 1.
DR Pfam; PF02739; 5_3_exonuc_N; 1.
DR Pfam; PF00476; DNA_pol_A; 1.
DR PRINTS; PR00868; DNAPOLI.
DR SMART; SM00474; 35EXOC; 1.
DR SMART; SM00475; 53EXOC; 1.
DR SMART; SM00278; HHH1; 1.
DR SMART; SM00279; HHH2; 1.
DR SMART; SM00482; POLAC; 1.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
DR TRANSFERASE; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydrolase; Exonuclease; DNA-binding; Complete proteome.
SQ SEQUENCE 904 AA; 98471 MW; 1C8E560FE5F74323 CRC64;

Query Match 29.2%; Score 50; DB 1; Length 904;
Best Local Similarity 42.4%; Pred. No. 47;
Matches 14; Conservative 1; Mismatches 8; Indels 10; Gaps 1;

QY 10 LAARAGPNGIEG-----PTLROWLAARA 32
DB 302 LAAGGPVEGDFVGGALAPGTVROWLAHA 334

RESULT 11
CBBR_XANFL
ID CBBR_XANFL STANDARD; PRT; 333 AA.
AC P25545;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Rubisco operon transcriptional regulator.
GN CBBR OR CFXO.
OS Xanthobacter flavus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hyphomicrobium group; Xanthobacter.
OX NCBI_TaxID=281;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H4-14;
RX MEDLINE=94012468; PubMed=8407781;

```


EMBL; AF038633; AAC24202.1; -.
MGD; MGI:1276123; Ucn.
InterPro: IPR000187; CRF.
InterPro: IPR003620; Urocortin_CRF.
Pfam: PF00473; CRF; 1.
ProDom; PD005970; Urocortin_CRF; 1.
SMART; SM00039; CRF; 1.

EMBL; M60848; AAA67901.1; -
EMBL; X51962; CAA36222.1; -
PIR; S06533; S06533.
DNA recombination; DNA-binding.
SEQUENCE 72 AA; 8635 MW; 0E6A4843503344A CRC64;

```

Query Match          28.1%; Score 48; DB 1; Length 72;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches          9; Conservative          7; Mismatches          6; Indels          2; Gaps          1;

y      5 TLQWLAAAGPAGNGIEGPTLRQWL 28
      ||::|||::|::|::|::|::|::|::|
b      4 TLQEWNAQRQRRPSLE--TVRRWV 25

RESULT 18
XIS_LAMB
XIS_LAMB STANDARD; PRT; 72 AA.
C      P03699;
C      21-JUL-1986 (Rel. 01, Created)
T      21-JUL-1986 (Rel. 01, Last sequence update)
T      01-AUG-1992 (Rel. 23, Last annotation update)
E      Excisionase.
D      XIS.
S      Bacteriophage lambda.
S      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
C      Lambda phage group.
C      NCBI_TaxID=10710;
X      [1]
X      SEQUENCE FROM N.A.
X      MEDLINE=83189071; PubMed=6221115;
X      Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.:
X      "Nucleotide sequence of bacteriophage lambda DNA.";
X      J. Mol. Biol. 162:729-773(1982).
X      [2]
X      SEQUENCE FROM N.A.
X      MEDLINE=81053845; PubMed=6253947;
X      Davies R.W.;
X      "DNA sequence of the int-xis-pi region of the bacteriophage lambda;
X      overlap of the int and xis genes.";
X      Nucleic Acids Res. 8:1765-1782(1980).
X      [3]
X      SEQUENCE FROM N.A.
X      MEDLINE=80234646; PubMed=6446713;
X      Hoess R.H., Foeller C., Bidwell K., Landy A.;
X      "Site-specific recombination functions of bacteriophage lambda: DNA
X      sequence of regulatory regions and overlapping structural genes for
X      int and xis.";
X      Proc. Natl. Acad. Sci. U.S.A. 77:2482-2486(1980).
X      -!- FUNCTION: EXCISIONASE AND INTEGRASE ARE NECESSARY FOR THE EXCISION
X      OF PROPHAGE FROM THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION AT
X      THE ATT SITE.
X      -----
X      This SWISS-PROT entry is copyright. It is produced through a collaboration
X      between the Swiss Institute of Bioinformatics and the EMBL outstation -
X      the European Bioinformatics Institute. There are no restrictions on its
X      use by non-profit institutions as long as its content is in no way
X      modified and this statement is not removed. Usage by and for commercial
X      entities requires a license agreement (See http://www.isb-sib.ch/announce/
X      or send an email to license@isb-sib.ch).
X      -----
X      EMBL; J02459; AAA96563.1; -.
X      PIR; A04321; RSBPXL
X      DNA recombination; DNA-binding.
X      SEQUENCE 72 AA; 8605 MW; 0E6A4843502200AA CRC64;

Query Match          28.1%; Score 48; DB 1; Length 72;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches          9; Conservative          7; Mismatches          6; Indels          2; Gaps          1;

y      5 TLQWLAAAGPAGNGIEGPTLRQWL 28
      ||::|||::|::|::|::|::|::|
b      4 TLQEWNAQRQRRPSLE--TVRRWV 25

RESULT 19
XIS_LAMB
XIS_LAMB STANDARD; PRT; 72 AA.
C      P03699;
C      21-JUL-1986 (Rel. 01, Created)
T      21-JUL-1986 (Rel. 01, Last sequence update)
T      01-AUG-1992 (Rel. 23, Last annotation update)
E      Excisionase.
D      XIS.
S      Bacteriophage lambda.
S      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
C      Lambda phage group.
C      NCBI_TaxID=10710;
X      [1]
X      SEQUENCE FROM N.A.
X      MEDLINE=83189071; PubMed=6221115;
X      Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.:
X      "Nucleotide sequence of bacteriophage lambda DNA.";
X      J. Mol. Biol. 162:729-773(1982).
X      [2]
X      SEQUENCE FROM N.A.
X      MEDLINE=81053845; PubMed=6253947;
X      Davies R.W.;
X      "DNA sequence of the int-xis-pi region of the bacteriophage lambda;
X      overlap of the int and xis genes.";
X      Nucleic Acids Res. 8:1765-1782(1980).
X      [3]
X      SEQUENCE FROM N.A.
X      MEDLINE=80234646; PubMed=6446713;
X      Hoess R.H., Foeller C., Bidwell K., Landy A.;
X      "Site-specific recombination functions of bacteriophage lambda: DNA
X      sequence of regulatory regions and overlapping structural genes for
X      int and xis.";
X      Proc. Natl. Acad. Sci. U.S.A. 77:2482-2486(1980).
X      -!- FUNCTION: EXCISIONASE AND INTEGRASE ARE NECESSARY FOR THE EXCISION
X      OF PROPHAGE FROM THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION AT
X      THE ATT SITE.
X      -----
X      This SWISS-PROT entry is copyright. It is produced through a collaboration
X      between the Swiss Institute of Bioinformatics and the EMBL outstation -
X      the European Bioinformatics Institute. There are no restrictions on its
X      use by non-profit institutions as long as its content is in no way
X      modified and this statement is not removed. Usage by and for commercial
X      entities requires a license agreement (See http://www.isb-sib.ch/announce/
X      or send an email to license@isb-sib.ch).
X      -----
X      EMBL; J02459; AAA96563.1; -.
X      PIR; A04321; RSBPXL
X      DNA recombination; DNA-binding.
X      SEQUENCE 72 AA; 8605 MW; 0E6A4843502200AA CRC64;

Query Match          28.1%; Score 48; DB 1; Length 72;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches          9; Conservative          7; Mismatches          6; Indels          2; Gaps          1;

y      5 TLQWLAAAGPAGNGIEGPTLRQWL 28
      ||::|||::|::|::|::|::|::|
b      4 TLQEWNAQRQRRPSLE--TVRRWV 25

RESULT 19
XIS_LAMB
XIS_LAMB STANDARD; PRT; 72 AA.
C      P03699;
C      21-JUL-1986 (Rel. 01, Created)
T      21-JUL-1986 (Rel. 01, Last sequence update)
T      01-AUG-1992 (Rel. 23, Last annotation update)
E      Excisionase.
D      XIS.
S      Bacteriophage lambda.
S      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
C      Lambda phage group.
C      NCBI_TaxID=10710;
X      [1]
X      SEQUENCE FROM N.A.
X      MEDLINE=83189071; PubMed=6221115;
X      Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.:
X      "Nucleotide sequence of bacteriophage lambda DNA.";
X      J. Mol. Biol. 162:729-773(1982).
X      [2]
X      SEQUENCE FROM N.A.
X      MEDLINE=81053845; PubMed=6253947;
X      Davies R.W.;
X      "DNA sequence of the int-xis-pi region of the bacteriophage lambda;
X      overlap of the int and xis genes.";
X      Nucleic Acids Res. 8:1765-1782(1980).
X      [3]
X      SEQUENCE FROM N.A.
X      MEDLINE=80234646; PubMed=6446713;
X      Hoess R.H., Foeller C., Bidwell K., Landy A.;
X      "Site-specific recombination functions of bacteriophage lambda: DNA
X      sequence of regulatory regions and overlapping structural genes for
X      int and xis.";
X      Proc. Natl. Acad. Sci. U.S.A. 77:2482-2486(1980).
X      -!- FUNCTION: EXCISIONASE AND INTEGRASE ARE NECESSARY FOR THE EXCISION
X      OF PROPHAGE FROM THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION AT
X      THE ATT SITE.
X      -----
X      This SWISS-PROT entry is copyright. It is produced through a collaboration
X      between the Swiss Institute of Bioinformatics and the EMBL outstation -
X      the European Bioinformatics Institute. There are no restrictions on its
X      use by non-profit institutions as long as its content is in no way
X      modified and this statement is not removed. Usage by and for commercial
X      entities requires a license agreement (See http://www.isb-sib.ch/announce/
X      or send an email to license@isb-sib.ch).
X      -----
X      EMBL; J02459; AAA96563.1; -.
X      PIR; A04321; RSBPXL
X      DNA recombination; DNA-binding.
X      SEQUENCE 72 AA; 8605 MW; 0E6A4843502200AA CRC64;

Query Match          28.1%; Score 48; DB 1; Length 72;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches          9; Conservative          7; Mismatches          6; Indels          2; Gaps          1;

y      5 TLQWLAAAGPAGNGIEGPTLRQWL 28
      ||::|||::|::|::|::|::|::|
b      4 TLQEWNAQRQRRPSLE--TVRRWV 25

RESULT 19
XIS_LAMB
XIS_LAMB STANDARD; PRT; 72 AA.
C      P03699;
C      21-JUL-1986 (Rel. 01, Created)
T      21-JUL-1986 (Rel. 01, Last sequence update)
T      01-AUG-1992 (Rel. 23, Last annotation update)
E      Excisionase.
D      XIS.
S      Bacteriophage lambda.
S      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
C      Lambda phage group.
C      NCBI_TaxID=10710;
X      [1]
X      SEQUENCE FROM N.A.
X      MEDLINE=83189071; PubMed=6221115;
X      Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.:
X      "Nucleotide sequence of bacteriophage lambda DNA.";
X      J. Mol. Biol. 162:729-773(1982).
X      [2]
X      SEQUENCE FROM N.A.
X      MEDLINE=81053845; PubMed=6253947;
X      Davies R.W.;
X      "DNA sequence of the int-xis-pi region of the bacteriophage lambda;
X      overlap of the int and xis genes.";
X      Nucleic Acids Res. 8:1765-1782(1980).
X      [3]
X      SEQUENCE FROM N.A.
X      MEDLINE=80234646; PubMed=6446713;
X      Hoess R.H., Foeller C., Bidwell K., Landy A.;
X      "Site-specific recombination functions of bacteriophage lambda: DNA
X      sequence of regulatory regions and overlapping structural genes for
X      int and xis.";
X      Proc. Natl. Acad. Sci. U.S.A. 77:2482-2486(1980).
X      -!- FUNCTION: EXCISIONASE AND INTEGRASE ARE NECESSARY FOR THE EXCISION
X      OF PROPHAGE FROM THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION AT
X      THE ATT SITE.
X      -----
X      This SWISS-PROT entry is copyright. It is produced through a collaboration
X      between the Swiss Institute of Bioinformatics and the EMBL outstation -
X      the European Bioinformatics Institute. There are no restrictions on its
X      use by non-profit institutions as long as its content is in no way
X      modified and this statement is not removed. Usage by and for commercial
X      entities requires a license agreement (See http://www.isb-sib.ch/announce/
X      or send an email to license@isb-sib.ch).
X      -----
X      EMBL; J02459; AAA96563.1; -.
X      PIR; A04321; RSBPXL
X      DNA recombination; DNA-binding.
X      SEQUENCE 72 AA; 8605 MW; 0E6A4843502200AA CRC64;

Query Match          28.1%; Score 48; DB 1; Length 72;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches          9; Conservative          7; Mismatches          6; Indels          2; Gaps          1;

y      5 TLQWLAAAGPAGNGIEGPTLRQWL 28
      ||::|||::|::|::|::|::|::|
b      4 TLQEWNAQRQRRPSLE--TVRRWV 25

RESULT 19
XIS_LAMB
XIS_LAMB STANDARD; PRT; 72 AA.
C      P03699;
C      21-JUL-1986 (Rel. 01, Created)
T      21-JUL-1986 (Rel. 01, Last sequence update)
T      01-AUG-1992 (Rel. 23, Last annotation update)
E     
```

RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 CC -!- FUNCTION: Participates in site-specific recombination. Acts by
 CC catalyzing the cutting and rejoining of the recombining DNA
 CC molecules. Acts jointly with XerD (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z97369; CAB10656.1; ALT_INIT.
 DR EMBL; AL583922; CAC30551.1; -.
 DR Leprosia; ML1600; -.
 DR InterPro; IPR002104; Phage_integrase.
 DR Pfam; PF00589; Phage_integrase; 1.
 KW DNA recombination; DNA integration; Complete proteome.
 FT ACT_SITE 278 TRANSIENT COVALENT LINKAGE TO DNA DURING
 FT ACT_SITE 278 STRAND CLEAVAGE AND REJOINING (BY
 FT SIMILARITY).
 SQ SEQUENCE 297 AA; 32180 MW; E70FA43F15286053 CRC64;
 Query Match 28.1%; Score 48; DB 1; Length 297;
 Best Local Similarity 37.9%; Pred. No. 28;
 Matches 11; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
 QY 4 PTLROWLAARAGPNCIEGPTLROWLAARA 32
 DB 49 PVLRSWLATAAGAGAARTTLARRISAVKA 77
 RESULT 21
 ODPB_BACSU STANDARD; PRT; 370 AA.
 AC P21881; Q59227;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pyruvate dehydrogenase E1 component, alpha subunit (EC 1.2.4.1) (S
 DE complex, 42 kDa subunit) (Vegetative protein 220) (VEG220).
 GN PDHA OR ACEA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=90368558; PubMed=1697575;
 RA Hemila H., Palva A., Paulin L., Arvidsson S., Palva I.;
 RT "Secretory S complex of Bacillus subtilis: sequence analysis and
 RT identity to pyruvate dehydrogenase.";
 RL J. Bacteriol. 172:5052-5063(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=97124187; PubMed=8969500;
 RA Winters P., Caldwell R., Enfield L., Ferrari E.;
 RT "The ampS-nprE (124 degrees-127 degrees) region of the Bacillus
 RT subtilis 168 chromosome: sequencing of a 27 kb segment and
 RT identification of several genes in the area.";
 RL Microbiology 142:3033-3037(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX Caldwell R.M., Ferrari E.;
 RA "Sequence analysis of the mobA-ampS region of the Bacillus subtilis
 RT chromosome.";
 RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

[4]
 RN SEQUENCE OF 1-15.
 RP STRAIN=ISS8;
 RX MEDLINE=97443988; PubMed=9298659;
 RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
 RA Hecker M.;
 RT "First steps from a two-dimensional protein index towards a response-
 RT regulation map for Bacillus subtilis.";
 RL Electrophoresis 18:1451-1463(1997).
 CC -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
 CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
 CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
 CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
 CC (E3).
 CC -!- FUNCTION: THE B.SUBTILIS PDH COMPLEX POSSESSES ALSO BRANCHED-CHAIN
 CC 2-OXOACID DEHYDROGENASE (BCDH) ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: Pyruvate + lipoamide = S-
 CC acetyldihydrolipoamide + CO(2).
 CC -!- COFACTOR: THIAMINE PYROPHOSPHATE.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M57435; AAA62681.1; -.
 DR EMBL; AF012285; AAC24932.1; -.
 DR EMBL; Z99111; CAB13331.1; -.
 DR PIR; B36718; DEBSEA.
 DR HSP; P09060; IQS0.
 DR Subtilisin; BG10207; pdhA.
 DR InterPro; IPR001017; E1_dh.
 DR Pfam; PF00676; E1_dehydrog; 1.
 KW Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
 KW Complete proteome.
 FT INIT_MET 0
 FT CONFLICT 178 178 A -> R (IN REF. 1).
 SQ SEQUENCE 370 AA; 41417 MW; 3183EB8881E1BD6D CRC64;
 Query Match 28.1%; Score 48; DB 1; Length 370;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 12; Conservative 3; Mismatches 7; Indels 2; Gaps 1;
 QY 2 EGPTLROWLAARAGPNCIEG--PT 23
 DB 258 EGPTLTTLTFRYGPHTMAGDDPT 281
 RESULT 22
 CA21_HUMAN
 ID CA21_HUMAN STANDARD; PRT; 1366 AA.
 AC P08123; P02464; Q9UEB6; Q9UPH0;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(I) chain precursor.
 GN COL1A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88058962; PubMed=2824475;
 RA de Wet W.J., Bernard M.P., Benson-Chanda V., Chu M.-L., Dickson L.A.,
 RA Weil D., Ramirez F.;
 RT "Organization of the human pro-alpha 2(I) collagen gene.";
 RT J. Biol. Chem. 262:16032-16036(1987).
 RN [2]

RP SEQUENCE FROM N.A.
RA Korkko J.M., Earley J.J., Ala-Kokko L., Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by CSGE and DNA sequencing in
RT 14 patients with mild OI (type I). Identification of common sequences
RT for null allele mutations.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88339824; PubMed=3421913;
RA Kuivaniemi H., Tromp G., Chu M.-L., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 2(I)
RT chain of human type I procollagen. Comparison with the chicken gene
RT confirms unusual patterns of gene conservation.";
RL Biochem. J. 252:633-640(1988).
RN [4]
RN SEQUENCE OF 181-1366 FROM N.A.
RA Kalicki J., Wamsley P., Gibson A.;
RN Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 623-1366 FROM N.A.
RA Bernard M.P., Myers J.C., Chu M.-L., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Structure of a cDNA for the pro alpha 2 chain of human type I
RT procollagen. Comparison with chick cDNA for pro alpha 2(I) identifies
RT structurally conserved features of the protein and the gene.";
RL Biochemistry 22:1139-1145(1983).
RN [6]
RN SEQUENCE OF 80-96.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [7]
RN SEQUENCE OF 417-447.
RC TISSUE=Skin;
RX MEDLINE=75008198; PubMed=4412529;
RA Fietzek P.P., Furtmayr H., Kuehn K.;
RT "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
RT and pig-skin collagen.";
RL Eur. J. Biochem. 47:257-261(1974).
RN [8]
RN SEQUENCE OF 145-198 FROM N.A.
RX MEDLINE=88298792; PubMed=3403536;
RA Kuivaniemi H., Sabol C., Tromp G., Sippola-Thiele M., Prockop D.J.;
RT "A 19-base pair deletion in the pro-alpha 2(I) gene of type I
RT procollagen that causes in-frame RNA splicing from exon 10 to exon 12
RT in a proband with atypical osteogenesis imperfecta and in his
RT asymptomatic mother.";
RL J. Biol. Chem. 263:11407-11413(1988).
RN [9]
RN SEQUENCE OF 960-1351 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=90304220; PubMed=2364107;
RA Maekelae J.K., Vuorio T., Vuorio E.;
RT "Growth-dependent modulation of type I collagen production and mRNA
RT levels in cultured human skin fibroblasts.";
RL Biochim. Biophys. Acta 1049:171-176(1990).
RN [10]
RN REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RN REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-

RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RN REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis Imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RN REVIEW ON OI VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalgleish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RN VARIANT EDS-VII-A2.
RX MEDLINE=88059013; PubMed=3680255;
RA Wirtz M.K., Glanville R.W., Steinmann B., Rao V.H., Hollister D.W.;
RT "Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids
RT comprising the N-telopeptide region of a pro-alpha 2(I) chain.";
RL J. Biol. Chem. 262:16376-16385(1987).
RN [15]
RN SEQUENCE OF 1090-1107 FROM N.A., AND VARIANT OI-IV ARG-1102.
RX MEDLINE=88227975; PubMed=2897363;
RA Wenstrup R.J., Cohn D.H., Cohen T., Byers P.H.;
RT "Arginine for glycine substitution in the triple-helical domain of
RT the products of one alpha 2(I) collagen allele (COL1A2) produces the
RT osteogenesis imperfecta type IV phenotype.";
RL J. Biol. Chem. 263:7734-7740(1988).
RN [16]
RN VARIANT OI-II ASP-997.
RX MEDLINE=89123407; PubMed=2914942;
RA Baldwin C.T., Constantinou C., Dumars K.W., Prockop D.J.;
RT "A single base mutation that converts glycine 907 of the alpha 2(I)
RT chain of type I procollagen to aspartate in a lethal variant of
RT osteogenesis imperfecta. The single amino acid substitution near the
RT carboxyl terminus destabilizes the whole triple helix.";
RL J. Biol. Chem. 264:3002-3006(1989).
RN [17]
RN VARIANT OI-II SER-955.
RX MEDLINE=89380165; PubMed=2777764;
RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT "Characterization of point mutations in the collagen COL1A1 and
RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL J. Biol. Chem. 264:15809-15812(1989).
RN [18]
RN VARIANT OI-II CYS-877.
RA Firtala A., Westerhausen A., Morris G.M., Rooney J.E., Prockop D.J.;
RT "Two cysteine substitutions in the type I procollagen genes (COL1A1
RT and COL1A2) that cause lethal osteogenesis imperfecta. The location
RT of glycine substitutions does not in any simple way predict their
RT effects on protein function or phenotype.";
RL Am. J. Hum. Genet. 47:A216-A216(1990).
RN [19]
RN VARIANT EDS-VII-A2.
RX MEDLINE=90368825; PubMed=2394758;
RA Weil D., D'Alessio M., Ramirez F., Eyre D.R.;
RT "Structural and functional characterization of a splicing mutation in
RT the pro-alpha 2(I) collagen gene of an Ehlers-Danlos type VII
RT patient.";
RL J. Biol. Chem. 265:16007-16011(1990).
RN [20]
RN VARIANTS OI-IV VAL-676.
RX MEDLINE=91291136; PubMed=2064612;
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN [21]
RN VARIANTS OI CYS-349 AND CYS-736.
RX MEDLINE=91115889; PubMed=1990009;

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphoribosylformylglycinamide synthase (EC 6.3.5.3) (FGAM
 DE synthase) (FGAMS) (Formylglycinamide ribotide amidotransferase)
 DE (FGARAT) (Formylglycinamide ribotide synthetase).
 GN PFAS OR KIA0361.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20018191; PubMed=10548741;
 RA Patterson D., Bleskan J., Gardiner K., Bowersox J.;
 RA "Human phosphoribosylformylglycinamide amidotransferase (FGARAT):
 RT regional mapping, complete coding sequence, isolation of a functional
 RT genomic clone, and DNA sequence analysis.";
 RL Gene 239:381-391(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide + L-
 CC glutamine + H₂O -> ADP + phosphate + 5'-
 CC phosphoribosylformylglycinamide + L-glutamate.
 CC -!- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FGAMS
 CC FAMILY.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO TYPE-1 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB002359; BAA20816.1; ALT_INIT.
 DR MIM; 602133;
 DR InterPro: IPR000728; AIRS_related.
 DR Pfam; PF00586; AIRS; 1.
 DR Pfam; PF02769; AIRS_C; 2.
 KW Purine biosynthesis; Ligase; ATP-binding; Glutamine amidotransferase.
 FT NP_BIND 322 333
 FT ACT_SITE 1158 1158 GATASE (BY SIMILARITY).
 SQ SEQUENCE 1338 AA; 144663 MW; 9741F8EDB8E1FE8 CRC64;
 Query Match 27.5%; Score 47; DB 1; Length 1338;
 Best Local Similarity 35.7%; Pred. NO. 1.7e+02;
 Matches 10; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
 QY 8 QWLAARAGP-----NGIEGPTLRQWLA 29
 Db 1205 RWASRVGPGPALMURMEGAVLPWNSA 1232
 RESULT 28
 ID CA21_MOUSE STANDARD; PRT: 1372 AA.
 AC Q01149;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 2(I) chain precursor.
 GN COL1A2 OR COLA2.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Calvaria;
 RA MEDLINE=92372043; PubMed=1505972;
 RA Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;
 RA "Sequence analysis of a full-length cDNA for the murine pro alpha
 RT 2(I) collagen chain: comparison of the derived primary structure with
 RT human pro alpha 2(I) collagen.";
 RL Genomics 13:1345-1346(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Breast tumor;
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-110 FROM N.A.
 RX TISSUE=Calvaria;
 RA MEDLINE=92084969; PubMed=1748823;
 RA Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,
 RA Wenstrup R.J.;
 RA "Construction of a full-length murine pro alpha 2(I) collagen cDNA by
 RT the polymerase chain reaction.";
 RL J. Invest. Dermatol. 97:980-984(1991).
 RN [4]
 RP SEQUENCE OF 1-23 FROM N.A.
 RX MEDLINE=87289650; PubMed=3039494;
 RA Rossi P., de Crombrughe B.;
 RA "Identification of a cell-specific transcriptional enhancer in the
 RT first intron of the mouse alpha 2 (type I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987).
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (FIBRILLAR FORMING COLLAGEN).
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYAPATITE.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X58251; CAA41205.1; -;
 DR EMBL; BC007158; AA07158.1; -;
 DR EMBL; K01832; AAA37331.1; -;
 DR PIR; A43291; A43291.
 DR MGD; MGI:88468; Colla2.
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib_collagen_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 85 AMINO-TERMINAL PROPEPTIDE
 FT (BY SIMILARITY).
 FT CHAIN 86 1108 COLLAGEN ALPHA 2(I) CHAIN.
 FT PROPEP 1109 1372 CARBOXYL-TERMINAL PROPEPTIDE
 FT (BY SIMILARITY).
 FT MOD_RES 86 86 PYRROLIDONE CARBOXYLIC ACID (BY
 FT SIMILARITY).
 FT MOD_RES 90 90 CONVERTED TO AN ALDEHYDE GROUP THAT IS

```

FT      CARBOHYD 1273 1273 INVOLVED IN CROSS-LINKING
FT      CONFLICT 15 15 (BY SIMILARITY).
FT      CONFLICT 1167 1167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CONFLICT 1167 1167 V -> A (IN REF. 4).
SQ      SEQUENCE 1372 AA; 129557 MW; 0D17DF3D6C1452D1 CRC64;
      Query Match 27.5%; Score 47; DB 1; Length 1372;
      Best Local Similarity 50.0%; Pred. No. 1.7e+02;
      Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY      1 IEQPTLRQWLAAAGPNGIEGTP 22
      IIII I :IIII II
DB      758 IVGPTGSVGAAGSPGPGPPG 779

RESULT 29
IE18_PRIVK STANDARD; PRT; 1446 AA.
AC      P33479;
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DT      01-FEB-1994 (Rel. 28, Last annotation update)
DE      Immediate-early protein IE180.
GN      IE.
OS      Pseudorabies virus (strain Kaplan) (PRV).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Alphaherpesvirinae; Varicellovirus.
OX      NCBI_TaxID=33703;
RN      [1]
RP      MEDLINE=91021039; PubMed=21711211;
RA      Vleck C., Kozmik Z., Paces V., Schirm S., Schwyzler M.;
RT      "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT      oriented open reading frame: characterization of their promoter and
RT      enhancer regions.";
RL      Virology 179:365-377(1990).
CC      -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC      OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC      OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC      -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC      -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC      PHOSPHORYLATION.
CC      -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M34651; AAA47470.1; -
DR      PIR; A5344; A5344.
KW      Early protein; Transcription regulation; Trans-acting factor;
FT      DNA-binding; Phosphorylation; Nuclear protein.
FT      DOMAIN 347 354 POLY-SER.
FT      DOMAIN 379 397 POLY-SER.
SQ      SEQUENCE 1446 AA; 148640 MW; 81F43A3DE3DDA068 CRC64;

Query Match 27.5%; Score 47; DB 1; Length 1446;
Best Local Similarity 44.0%; Pred. No. 1.8e+02;
Matches 11; Conservative 2; Mismatches 10; Indels 2; Gaps 1;

QY      3 GPTL--RWLAARAGPNGIEGPTLR 25
      II: I:I II I I II I
DB      182 GPSAAPRRWSPARGDPVGEFGPAAR 206

RESULT 30
IE18_PRIVF STANDARD; PRT; 1461 AA.
ID      IE18_PRIVF

```

```

AC      P11675;
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      01-FEB-1994 (Rel. 28, Last annotation update)
DE      Immediate-early protein IE180.
GN      IE.
OS      Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Alphaherpesvirinae; Varicellovirus.
OX      NCBI_TaxID=31523;
RN      [2]
RP      SEQUENCE FROM N.A.
RP      MEDLINE=89315207; PubMed=2546124;
RA      Cheung A.K.;
RT      "DNA nucleotide sequence analysis of the immediate-early gene of
RT      pseudorabies virus.";
RL      Nucleic Acids Res. 17:4637-4646(1989).
RN      [2]
RP      REVISIONS.
RA      Cheung A.K.;
RL      Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC      OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC      OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC      -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC      -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC      PHOSPHORYLATION.
CC      -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X15120; CAA33214.1; -
DR      PIR; S04713; EDBEIF.
KW      Early protein; Transcription regulation; Trans-acting factor;
KW      DNA-binding; Phosphorylation; Nuclear protein.
FT      DOMAIN 390 405 POLY-SER.
FT      DOMAIN 958 966 POLY-SER.
SQ      SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 27.5%; Score 47; DB 1; Length 1461;
Best Local Similarity 44.0%; Pred. No. 1.9e+02;
Matches 11; Conservative 2; Mismatches 10; Indels 2; Gaps 1;

QY      3 GPTL--RWLAARAGPNGIEGPTLR 25
      II: I:I II I I II I
DB      190 GPSAAPRRWSPARGDPVGEFGPAAR 214

Search completed: October 9, 2002, 09:00:08
Job time : 4.90535 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 11.466 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-23

Perfect score: 171

Sequence: 1 IEPTLRQLWLAARAGPNEGPTLRQLWLAARA 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000.

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL19:*
- 2: sp_archaea:*
- 3: sp_bacteria:*
- 4: sp_fungi:*
- 5: sp_human:*
- 6: sp_invertebrate:*
- 7: sp_mammal:*
- 8: sp_mhc:*
- 9: sp_organelle:*
- 10: sp_phase:*
- 11: sp_plant:*
- 12: sp_rodent:*
- 13: sp_virus:*
- 14: sp_vertebrate:*
- 15: sp_unclassified:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	36.8	683	16	083436
2	62	36.3	607	2	Q9L8D4
3	60	35.1	509	2	Q9S5E5
4	58.5	34.2	869	5	Q9VZ82
5	56	32.7	361	16	Q9ABC7
6	55.5	32.5	1744	3	Q9A192
7	55	32.2	420	2	P97011
8	54	31.6	1095	16	Q9I304
9	53	31.0	305	2	Q9S0M9
10	53	31.0	326	16	Q9RTE6
11	52.5	30.7	814	4	Q96C78
12	52	30.4	396	2	Q9X7N5
13	52	30.4	902	5	O31611
14	52	30.4	967	2	Q9KZD5
15	52	30.4	1349	2	Q9L096
16	51.5	30.1	371	16	Q9I477

17	51	29.8	281	17	Q9YDQ0
18	51	29.8	306	16	005576
19	51	29.8	322	2	Q9RK51
20	51	29.8	381	2	Q9X757
21	51	29.8	589	5	Q18756
22	51	29.8	600	16	Q9HXJ8
23	51	29.8	719	16	Q922H9
24	51	29.8	1820	13	Q9I907
25	50.5	29.5	526	16	Q981N1
26	50.5	29.5	1460	5	Q9GY79
27	50	29.2	250	2	Q93LY8
28	50	29.2	268	16	Q98LG1
29	50	29.2	351	16	Q9RWB0
30	50	29.2	384	2	Q9F2F9
31	50	29.2	400	2	Q9XDB0
32	50	29.2	403	2	Q9F5B8
33	50	29.2	410	16	Q9I1M2
34	50	29.2	472	5	017754
35	50	29.2	604	16	Q98P10
36	50	29.2	1272	4	Q9UGH1
37	50	29.2	1300	4	Q9BXA9
38	49.5	28.9	333	10	Q94LX0
39	49.5	28.9	336	11	Q9AV74
40	49	28.7	133	2	Q9AOH5
41	49	28.7	214	5	Q20968
42	49	28.7	249	2	Q9L3H3
43	49	28.7	250	10	Q9AS26
44	49	28.7	307	10	Q43416
45	49	28.7	319	2	Q9RKM5

ALIGNMENTS

RESULT 1

083436 ID 083436 PRELIMINARY; PRT; 683 AA.
AC 083436;
DT 01-NOV-1998 (TREMUREL. 08, Created)
DT 01-NOV-1998 (TREMUREL. 08, Last sequence update)
DE 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0421.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G., Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J., Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T., McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis spirochete.";
RT Spirochete.";
RL Science 281:375-388(1998).
EMBL; AE001220; AAC65409.1;
TIGR; TP0421;
DR InterPro; IPR001258; NHL.
DR InterPro; IPR001440; TPR.
DR Pfam; PF01436; NHL; 4.
DR Pfam; PF00515; TPR; 1.
KW Complete proteome.
SQ SEQUENCE 683 AA; 74518 MW; F91407FA7094AD1 CRC64;

Query Match 36.8%; Score 63; DB 16; Length 683;
Best Local Similarity 46.4%; Pred. No. 4.6;
Matches 13; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Q9YDQ0 aeropyrum p
005576 mycobacteri
Q9RK51 streptomyce
Q9X757 klebsiella
Q18756 caenorhabdi
Q9HYJ8 pseudomonas
Q922H9 rhizobium m
Q9I907 pagrus majo
Q981N1 rhizobium l
Q9GY79 leishmania
Q93LY8 streptomyce
Q98LG1 rhizobium l
Q9RWB0 deinococcus
Q9F2F9 streptomyce
Q9XDB0 mycobacteri
Q9F5B8 agrobacteri
Q9I1M2 pseudomonas
017754 caenorhabdi
Q98P10 rhizobium l
Q9UGH1 homo sapien
Q9BXA9 homo sapien
Q94LX0 perilla fru
Q9AV74 mus musculu
Q9AOH5 achromobact
Q20968 caenorhabdi
Q9L3H3 rhizobium l
Q9AS26 oryza sativ
Q43416 carchrus ci
Q9RKM5 streptomyce

RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb streptomycetes coelicolin A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AB017438; BAA82701.1; -.
 DR EMBL; AL356592; CAB92204.1; -.
 KW DNA-binding.
 SQ SEQUENCE 509 AA; 54398 MW; 7BB074DAAE0F1867 CRC64;
 Query Match 35.1%; Score 60; DB 2; Length 509;
 Best Local Similarity 44.1%; Pred. No. 8.2;
 Matches 15; Conservative 4; Mismatches 11; Indels 4; Gaps 2;
 QY 1 TEGPTLROW---LAARAGPNGIE-GPTLRQWLAA 30
 DB 404 LAGPALRTWAVDLGLRDPDGRDLRLTLRTWIAA 437
 RESULT 4
 QV9Z82 PRELIMINARY; PRT; 869 AA.
 ID Q9VZ82;
 AC Q9VZ82;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG7479 PROTEIN.
 DE CG7479
 GN Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophilinae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003482; AAF47943.1; -;
 DR FlyBase: FBgn0035576; CG7479.
 DR InterPro: IPR002300; trna-synt_la.
 DR InterPro: IPR001412; trna-synt_l.
 DR InterPro: IPR002302; trna-synt_leu.
 DR Pfam: PF00133; trna-synt.1; 1.
 DR PRINTS; PR00985; TRNASYNTHLEU.
 DR PROSITE; PS00178; AA.TRNA.LIGASE.1; 1.
 SQ SEQUENCE 869 AA; 99299 MW; E87A1ECBBEBB27B67 CRC64;

Query Match 34.2%; Score 58.5; DB 5; Length 869;
 Best Local Similarity 40.6%; Pred. No. 22;
 Matches 13; Conservative 4; Mismatches 10; Indels 5; Gaps 1;

QY 1 IEPTLRQWLA-----ARAGPNEGPTLRQW 27
 Db 213 VEKLLRQWFTITSAYAKOLLDGLDPTLRQW 244
 :| ||||| : :||| ||||| |

RESULT 5

Q9ABC7
 ID Q9ABC7 PRELIMINARY; PRT; 361 AA.
 AC Q9ABC7
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CATION EFFLUX FAMILY PROTEIN.
 GN CC0303.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Eisen J., Heidelberg J.F., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Niernan W.C., Feidlyum T.V., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka R., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Ghinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005704; AAK22290.1; -;
 DR TIGR; CC0303; -;
 DR InterPro: IPR002524; Cation_efflux.
 DR InterPro: IPR002395; Kininogen.
 DR Pfam: PF01545; Cation_efflux; 1.
 DR PRINTS; PR00334; KININOGEN.
 KW Complete proteome.
 SQ SEQUENCE 361 AA; 38180 MW; 1A4F7F0A7C62EEB0 CRC64;

Query Match 32.7%; Score 56; DB 16; Length 361;
 Best Local Similarity 54.5%; Pred. No. 19;
 Matches 12; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 10 LAARAGPNEGPTLRQWLAAR 31
 Db 266 LALDATPRGIDTKVRDWAAR 287
 || ||||| : |||||

RESULT 6

O94192
 ID O94192 PRELIMINARY; PRT; 1744 AA.
 AC O94192;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE CHITIN SYNTHASE.
 GN CHS4.
 OS Paracoccidioides brasiliensis.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Onygenales; mitosporic Onygenales; Paracoccidioides.
 OX NCBI_TaxID=121759;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20210320; PubMed=10746225;
 RA Nino-Vega G.A., Munro C.A., San-Blas G., Gooday G.W., Gow N.A.;
 RT "Differential expression of chitin synthase genes during temperature-
 induced dimorphic transitions in *Paracoccidioides brasiliensis*."
 RL Med. Mycol. 38:31-39(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Nino-Vega G.A., San-Blas G.;
 RT "Sequence analysis of the CHS4 gene of *Paracoccidioides*
 RT *brasiliensis*."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF107624; AAD19613.2; -;
 DR InterPro: IPR002923; Chitin_synth.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR001173; Glycos transf. 2.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF03142; Chitin_synth_2; 1.
 DR Pfam: PF00063; myosin_head; 1.
 DR SMART; SM00242; MYSc; 1.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN.1.
 SQ SEQUENCE 1744 AA; 193777 MW; DB7622D0A69F0705 CRC64;

Query Match 32.5%; Score 55.5; DB 3; Length 1744;
 Best Local Similarity 51.7%; Pred. No. 1.le+02;
 Matches 15; Conservative 2; Mismatches 11; Indels 1; Gaps 1;

QY 5 TLROWL-AARAGPNEGPTLRQWLAAR 32
 Db 56 TVNTWLTAA SPNGEVGGTIDADLARRA 84
 | : || || |||| | : || ||

RESULT 7

P97011
 ID P97011 PRELIMINARY; PRT; 420 AA.
 AC P97011;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SORBITOL OXIDASE.
 GN SOX.
 OS Streptomyces sp.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1931;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H-7775;
 RA Hirada K., Eto T., Yoshioka I., Oda K.;
 RT "Cloning of a gene encoding a sorbitol oxidase from *Streptomyces* sp.
 RT H-7775."
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB000519; BAA19135.1; -;
 DR InterPro: IPR001575; Oxid_FAD_bind.
 DR Pfam: PF01565; FAD_binding_4; 1.
 SQ SEQUENCE 420 AA; 45181 MW; EF3189045CAF0649 CRC64;

Query Match 32.2%; Score 55; DB 2; Length 420;
 Best Local Similarity 37.9%; Pred. No. 29;
 Matches 11; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 3 GPTLRQWLAARAGPNEGPTLRQWLAAR 31
 Db 215 GPVGQWLKQRYGDEGARSVMPAEWLGR 243
 || : || | | : || ||

RT damage in *Deinococcus radiodurans*.";

SQ SEQUENCE 814 AA; 87717 MW; 683A8368AD30996B CRC64;
 Query Match 30.7%; Score 52.5; DB 4; Length 814;
 Best Local Similarity 44.8%; Pred. No. 1.2e+02;
 Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;
 QY 1 IEGPTLROWLAARAGPNGIEGPTLROWLA 29
 :||| | :||:| | :||:| | :||:| |
 Db 728 LKGPTC-QYRAAQSGPSRPPQRRALLA 755
 :||| | :||:| | :||:| | :||:| |
 RESULT 12
 Q9X7N5 PRELIMINARY; PRT; 396 AA.
 AC Q9X7N5;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN SC5F2A.12C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapaitte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL049587; CAB40679.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 396 AA; 41908 MW; BCB465197F3A3F6E CRC64;
 Query Match 30.4%; Score 52; DB 2; Length 396;
 Best Local Similarity 34.4%; Pred. No. 66;
 Matches 11; Conservative 5; Mismatches 10; Indels 6; Gaps 2;
 QY 4 PTLROWLAARAGPNGIEGPT----LROWLAAR 31
 :||| | :||:| | :||:| | :||:| |
 Db 293 PPARWLSGLRAGP--EGPSAERRAKSWFSVR 322
 :||| | :||:| | :||:| | :||:| |
 RESULT 13
 O16161 PRELIMINARY; PRT; 902 AA.
 AC O16161;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PRECOLLAGEN P PRECURSOR.
 GN PRECOL-P.
 OS Mytilus edulis (Blue mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
 OC Mytiloidea; Mytilidae; Mytilus.
 OX NCBI_TaxID=6550;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=FOOT;
 RX MEDLINE=97442537; PubMed=9295275;

RA Coyne K.J., Qin X.X., Waite J.H.;
 RT "Extensible collagen in mussel byssus: A natural block copolymer."
 RL Science 277:1830-1832(1997).
 DR EMBL; AF015539; AAB80719.1; -;
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 7.
 KW Signal; Collagen.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 902 COLLAGEN P.
 SQ SEQUENCE 902 AA; 78526 MW; D1EF09DEA2BD9EF3 CRC64;
 Query Match 30.4%; Score 52; DB 5; Length 902;
 Best Local Similarity 52.4%; Pred. No. 1.6e+02;
 Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 2 EGPTLROWLAARAGPNGIEGP 22
 :||| | :||:| | :||:| | :||:| |
 Db 527 KGPTGAQGPAGPAGPSGEQGP 547
 :||| | :||:| | :||:| | :||:| |
 RESULT 14
 Q9KZD5 PRELIMINARY; PRT; 967 AA.
 ID Q9KZD5;
 AC Q9KZD5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PROBABLE NADH DEHYDROGENASE I COMPLEX, SUBUNIT.
 GN SC6F7.07.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapaitte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 CC -|- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL; AL353870; CAB89015.1; -;
 DR InterPro; IPR001750; Oxidored_q1.
 DR Pfam; PF00361; Oxidored_q1; 1.
 KW NAD; Oxidoreductase; Transmembrane.
 SQ SEQUENCE 967 AA; 101770 MW; A727564D3EEE05AC CRC64;
 Query Match 30.4%; Score 52; DB 2; Length 967;
 Best Local Similarity 53.6%; Pred. No. 1.7e+02;
 Matches 15; Conservative 1; Mismatches 8; Indels 4; Gaps 2;
 QY 9 WLAAR---AG-PNGIEGPTLROWLAAR 32
 :||| | :||:| | :||:| | :||:| |
 Db 782 WRARYESAGLPTGDEGPPERTWLAASA 809
 :||| | :||:| | :||:| | :||:| |
 RESULT 15
 Q9L096 PRELIMINARY; PRT; 1349 AA.
 ID Q9L096
 AC Q9L096;

DR InterPro: IPR000205; NAD_binding.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 371 AA; 39174 MW; 016D60440BAD50D7 CRC64;

Query Match 30.1%; Score 51.5; DB 16; Length 371;
Best Local Similarity 31.0%; Pred. No. 72;
Matches 13; Conservative 7; Mismatches 9; Indels 13; Gaps 2;

QY 4 PTLROWLAARAGP-----NGIEGPTLR---OWLAARA 32
DB 145 PNAARWLLDQAGPRLRLRYAEVSEVDGSRRLADGRWLSAEA 186

RESULT 17
QYDQ0 PRELIMINARY; PRT; 281 AA.
AC QYDQ0
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 32.1 KDA PROTEIN APE0867.
GN APE0867.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1.
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79847.1; -.
DR InterPro: IPR000130; ZN_MTPeptidse.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 281 AA; 32123 MW; 09AC9AF6F92CB41E CRC64;

Query Match 29.8%; Score 51; DB 17; Length 281;
Best Local Similarity 34.4%; Pred. No. 62;
Matches 11; Conservative 7; Mismatches 6; Indels 8; Gaps 2;

QY 5 TLRQWLAARAGP-----GIEGPTLRQWLAAR 31
DB 12 SLRQWRS---PNRYDIPGVDSPEVGNWLES 40

RESULT 18
Q05576 PRELIMINARY; PRT; 306 AA.
AC Q05576
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GALU.
GN GALU OR RV0993 OR MTC1237.07.
OS Mycobacterium tuberculosis
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 145.7 KDA PROTEIN.
GN SCC24.21.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Brown S.P., Harris D.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL163003; CAB86115.1; -.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam; PF00069; pkinase; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00136; SUBTILASE_ASF; UNKNOWN_1.
KW ATP-binding; Hypothetical protein; Transferrase.
SQ SEQUENCE 1349 AA; 145671 MW; F0902A235D694B38 CRC64;

Query Match 30.4%; Score 52; DB 2; Length 1349;
Best Local Similarity 51.6%; Pred. No. 2.4e+02;
Matches 16; Conservative 3; Mismatches 4; Indels 8; Gaps 4;

QY 3 GPTLRQWLAARAGPNGIEG-PTLRQWLAARA 32
DB 285 GP---QW---RAGPDGVRWFATLRW-ARRA 308

RESULT 16
Q91477 PRELIMINARY; PRT; 371 AA.
AC Q91477
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PA1267.
GN PA1267.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huftnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen";
RL Nature 406:959-964(2000).
RA EMBL; AE004556; AAG04656.1; -.

Q91907	PRELIMINARY;	PRT; 1820 AA.
AC	Q91907;	
AD	Q91907;	
DT	01-OCT-2000 (TReMBLrel. 15, Created)	
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)	
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	
DE	PRO-ALPHA 1 TYPE V/XI COLLAGEN.	
DE	COLV/XI.AL	
GN	Pagrus major (Red sea bream) (Chrysophrys major).	
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoldei;	
OC	Sparidae; Pagrus.	
OX	NCBI_TaxID=143350;	
OX	NCBI_TaxID=143350;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE-21240220; PubMed=11342118;	
RX	Touhata K., Tanaka H., Yokoyama Y., Sakaguchi M., Toyohara H.;	
RA	"Structure of a full-length cDNA clone for the pro-1(V/XI) collagen	
RA	chain of red seabream.";	
RT	Biochim. Biophys. Acta 1517:323-326(2001).	
RL	EMBL; AB045975; BAB03287.1; -	
DR	InterPro: IPR000087; Collagen.	
DR	InterPro: IPR000885; Fib_collagen_C.	
DR	InterPro: IPR001791; Laminin_G.	
DR	InterPro: IPR001230; Prenyltn.	
DR	InterPro: IPR003129; TSPN.	
DR	Pfam; PF01410; COLFI; 1.	
DR	Pfam; PF02210; TSPN; 1.	
DR	ProDom; PD002078; Fib_collagen_C; 1.	
DR	SMART; SM00038; COLFI; 1.	
DR	SMART; SM00282; LamG; 1.	
DR	SMART; SM00210; TSPN; 1.	
DR	PROSITE; PS00294; PRENYLATION; UNKNOWN_1.	
KW	Collagen.	
SQ	SEQUENCE 1820 AA; 181678 MW; 46E45E8AF7AD3DAE CRC64;	

Best Local Similarity 37.98; Pred. NO. 4.1e+02;
Matches 11; Conservative 6; Mismatches 7; Indels 5; Gaps 1;

QY 4 PTLRQLAARAGP-----NGIEGPTLRQW 27
::: ||| ||| |:::|
Db 725 PSLATAAAAAAGPYKSQNHQTPSMRW 753

```

RESULT 27
Q93LY8
ID Q93LY8 PRELIMINARY; PRT; 250 AA.
AC Q93LY8;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PGAK (FRAGMENT).
OS PGAK.
OS Streptomyces sp. PGA64.
OC Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomyces.
OC NCBI_taxID=161235;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.

```

110 LASNSAAVQERGVAAPTRVVARLAGFLLRHIEWLAHA 148

Q98LGI
Q98LGI PRELIMINARY; PRT; 268 AA.
AC Q98LGI;
Q98LGI

01-OCT-2001 (TReMBLrel. 18, Last annotation update)
PROBABLE SHORT CHAIN DEHYDROGENASE.
MIL11036.
Rhizobium loti (Mesorhizobium loti).
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
NCBI_TaxID=381;
[1]

W	Q	Complete proteome.
SEQUENCE	268 AA;	27788 MW; 86698FFD04036653 CRC64;

RA Patallo E.P.;
RT "deoxysugar methylation during biosynthesis of the antitumor
RT polyketide elloramycin by Streptomyces olivaceus: characterization of
RT three methyltransferase genes.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Blanco G.;
RT "Identification of a sugar flexible glycosyltransferase from
RT Streptomyces olivaceus, the producer of the antitumor polyketide
RT elloramycin.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ300305; CAC16413.1; -;
DR EMBL; AJ309821; CAC32467.1; -;
DR InterPro; IPR000890; Acetate_kin.
DR PROSITE; PS01076; ACETATE_KINASE_2; UNKNOWN_1.
KW Transferase.
SQ SEQUENCE 384 AA; 39674 MW; A254F56B6ED12F2B CRC64;

Query Match 29.2%; Score 50; DB 2; Length 384;
Best Local Similarity 52.2%; Pred. No. 1.2e+02;
Matches 12; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 6 LRQWLAARAGPNGIEGPTLRQWL 28
Db 152 VRQLLAERLGPAGSEPPERYFL 174

Search completed: October 9, 2002, 09:03:03
Job time : 12.5494 secs

RA Patallo E.P.;
RT "deoxysugar methylation during biosynthesis of the antitumor
RT polyketide elloramycin by Streptomyces olivaceus: characterization of
RT three methyltransferase genes.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Blanco G.;
RT "Identification of a sugar flexible glycosyltransferase from
RT Streptomyces olivaceus, the producer of the antitumor polyketide
RT elloramycin.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ300305; CAC16413.1; -;
DR EMBL; AJ309821; CAC32467.1; -;
DR InterPro; IPR000890; Acetate_kin.
DR PROSITE; PS01076; ACETATE_KINASE_2; UNKNOWN_1.
KW Transferase.
SQ SEQUENCE 384 AA; 39674 MW; A254F56B6ED12F2B CRC64;

Query Match 29.2%; Score 50; DB 2; Length 384;
Best Local Similarity 52.2%; Pred. No. 1.2e+02;
Matches 12; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 6 LRQWLAARAGPNGIEGPTLRQWL 28
Db 152 VRQLLAERLGPAGSEPPERYFL 174

Search completed: October 9, 2002, 09:03:03
Job time : 12.5494 secs

RA Patallo E.P.;
RT "deoxysugar methylation during biosynthesis of the antitumor
RT polyketide elloramycin by Streptomyces olivaceus: characterization of
RT three methyltransferase genes.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Blanco G.;
RT "Identification of a sugar flexible glycosyltransferase from
RT Streptomyces olivaceus, the producer of the antitumor polyketide
RT elloramycin.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ300305; CAC16413.1; -;
DR EMBL; AJ309821; CAC32467.1; -;
DR InterPro; IPR000890; Acetate_kin.
DR PROSITE; PS01076; ACETATE_KINASE_2; UNKNOWN_1.
KW Transferase.
SQ SEQUENCE 384 AA; 39674 MW; A254F56B6ED12F2B CRC64;

Query Match 29.2%; Score 50; DB 2; Length 384;
Best Local Similarity 52.2%; Pred. No. 1.2e+02;
Matches 12; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 6 LRQWLAARAGPNGIEGPTLRQWL 28
Db 152 VRQLLAERLGPAGSEPPERYFL 174

Search completed: October 9, 2002, 09:03:03
Job time : 12.5494 secs

RA Patallo E.P.;
RT "deoxysugar methylation during biosynthesis of the antitumor
RT polyketide elloramycin by Streptomyces olivaceus: characterization of
RT three methyltransferase genes.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Blanco G.;
RT "Identification of a sugar flexible glycosyltransferase from
RT Streptomyces olivaceus, the producer of the antitumor polyketide
RT elloramycin.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ300305; CAC16413.1; -;
DR EMBL; AJ309821; CAC32467.1; -;
DR InterPro; IPR000890; Acetate_kin.
DR PROSITE; PS01076; ACETATE_KINASE_2; UNKNOWN_1.
KW Transferase.
SQ SEQUENCE 384 AA; 39674 MW; A254F56B6ED12F2B CRC64;

Query Match 29.2%; Score 50; DB 2; Length 384;
Best Local Similarity 52.2%; Pred. No. 1.2e+02;
Matches 12; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 6 LRQWLAARAGPNGIEGPTLRQWL 28
Db 152 VRQLLAERLGPAGSEPPERYFL 174

Search completed: October 9, 2002, 09:03:03
Job time : 12.5494 secs

Wed Oct 9 10:29:36 2002

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 16.1874 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-24
Perfect score: 194
Sequence: 1 TSPPTLQWLAARAGGGGGGIEGPTLQWLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	194	100.0	36	21	AA16963
2	194	100.0	36	21	AA17293
3	194	100.0	36	21	AA196525
4	194	100.0	41	21	AA196528
5	194	100.0	42	21	AA17281
6	194	100.0	42	21	AA17282
7	194	100.0	42	21	AA17308
8	194	100.0	42	21	AA196530
9	194	100.0	60	21	AA17311
10	194	100.0	269	21	AA16960
11	194	100.0	269	21	AA196531

12	190	97.9	268	21	AA16959	Fc-TMP-TMP protein
13	186	95.9	36	21	AA17301	TPO-mimetic peptid
14	186	95.9	36	21	AA196523	Thrombopoietin mim
15	185	95.4	36	21	AA17303	TPO-mimetic peptid
16	185	95.4	36	21	AA17307	TPO-mimetic peptid
17	185	95.4	36	21	AA196524	Thrombopoietin mim
18	183.5	94.6	37	21	AA17294	TPO-mimetic peptid
19	183	94.3	38	21	AA17295	TPO-mimetic peptid
20	182.5	94.1	39	21	AA17304	TPO-mimetic peptid
21	182.5	94.1	39	21	AA17305	TPO-mimetic peptid
22	182	93.8	36	21	AA17306	TPO-mimetic peptid
23	182	93.8	36	21	AA196526	Thrombopoietin mim
24	181	93.3	42	21	AA17296	TPO-mimetic peptid
25	177.5	91.5	35	21	AA17292	TPO-mimetic peptid
26	174	89.7	40	21	AA17302	TPO-mimetic peptid
27	171	88.1	34	21	AA17291	TPO-mimetic peptid
28	168	86.6	36	21	AA17298	TPO-mimetic peptid
29	168	86.6	36	21	AA17299	TPO-mimetic peptid
30	166	85.6	36	21	AA196521	Cyclic or linear t
31	166	85.6	36	21	AA17300	TPO-mimetic peptid
32	166	85.6	36	21	AA196522	Linear thrombopoie
33	164.5	84.8	33	21	AA17290	TPO-mimetic peptid
34	158	81.4	32	21	AA17289	TPO-mimetic peptid
35	151.5	78.1	31	21	AA17288	TPO-mimetic peptid
36	145	74.7	30	21	AA17287	Thrombopoietin mim
37	144	74.2	32	21	AA17297	Thrombopoietin mim
38	144	74.2	32	21	AA196520	TPO-mimetic peptid
39	144	74.2	34	21	AA196527	TPO-mimetic peptid
40	138.5	71.4	29	21	AA17286	TPO-mimetic peptid
41	132	68.0	28	21	AA17285	TPO-mimetic peptid
42	131.5	67.8	29	21	AA16970	TPO-mimetic peptid
43	129.5	66.8	31	21	AA16973	TPO-mimetic peptid
44	129.5	66.8	31	21	AA16974	TPO-mimetic peptid
45	125.5	64.7	29	21	AA16971	TPO-mimetic peptid

ALIGNMENTS

RESULT 1
AA16963
ID AA16963 standard; Protein: 36 AA.
AC AA16963;
DT 31-OCT-2000 (first entry)
DE TPO-mimetic peptide TMP-TMP SEQ ID NO:14.
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
OS Synthetic.
XX WO200024782-A2.
XX 04-MAY-2000.
XX 25-OCT-1999; 99WO-US25044.
XX 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX (AMGE-) AMGEN INC.
XX Feige U, Liu C, Cheestham J, Boone TC;
XX WPI; 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX PS Disclosure; Page 190; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 36 AA;
 Query Match 100.0%; Score 194; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 1.7e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAARA 36
 |||||
 Db 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAARA 36
 |||||
 RESULT 2
 AAB17293
 ID AAB17293 standard; Peptide; 36 AA.
 XX
 AC AAB17293;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:349.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 XX WO200024782-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 25-OCT-1999; 99WO-US25044.
 XX
 XX PR 23-OCT-1998; 98US-0105371.
 XX
 XX PR 22-OCT-1999; 99US-0428082.
 XX
 XX PA (AMGE-) AMGEN INC.
 XX
 XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX
 XX DR WPI; 2000-350702/30.
 XX
 XX PT Novel composition of matter comprising an Fc domain and

PT PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX PS Example 1; Page 318; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 36 AA;
 Query Match 100.0%; Score 194; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 1.7e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAARA 36
 |||||
 Db 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAARA 36
 |||||
 RESULT 3
 AAY96525
 ID AAY96525 standard; Peptide; 36 AA.
 XX
 AC AAY96525;
 XX
 DT 04-SEP-2000 (first entry)
 XX
 DE Thrombopoietin mimetic peptide compound 6.
 XX
 KW Thrombopoietin; mimetic; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT Peptide /note= "optionally linked to an Fc molecule"
 FT 1..14 /label= TMP_1
 FT Peptide 15..18 /label= linker
 FT Peptide 19..32 /label= TMP_2
 FT Modified-site 32
 FT /note= "optionally linked to an Fc molecule"
 XX
 XX WO200024770-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 22-OCT-1999; 99WO-US24834.
 XX
 XX PR 23-OCT-1998; 98US-0105348.
 XX
 XX PA (AMGE-) AMGEN INC.
 XX

XX WPI; 2000-350702/30.

DR Novel composition of matter comprising an Fc domain and

XX pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

PT

XX Disclosure; Page 313; 608pp; English.

PS

XX The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each

CC independently provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ Sequence 42 AA;

Query Match 100.0%; Score 194; DB 21; Length 42;

Best Local Similarity 100.0%; Pred. No. 2e-16;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

DB 7 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

RESULT 6

AAB17282

ID AAB17282 standard; Peptide; 42 AA.

XX

AC AAB17282;

XX

DT 31-OCT-2000 (first entry)

XX

DE TPO-mimetic peptide sequence SEQ ID NO:338.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

XX

OS Synthetic.

XX

PN WO200024782-A2.

XX

PD 04-MAY-2000.

XX

PF 25-OCT-1999; 99WO-US25044.

XX

PR 23-OCT-1998; 98US-0105371.

XX

PR 22-OCT-1999; 99US-0428082.

XX

XX (AMGE-) AMGEN INC.

PA

XX Feige U, Liu C, Cheetham J, Boone TC;

PI

XX WPI; 2000-350702/30.

DR

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX Disclosure; Page 313; 608pp; English.

PS

XX The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:

CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,

CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4

CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each

CC independently linkers; and a, b, c, d, e, and f = are each independently

CC provided that at least 1 of a and b is 1. The composition can

CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX

SQ Sequence 42 AA;

Query Match 100.0%; Score 194; DB 21; Length 42;

Best Local Similarity 100.0%; Pred. No. 2e-16;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

DB 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 7

AAB17308

ID AAB17308 standard; Peptide; 42 AA.

XX

AC AAB17308;

XX

DT 31-OCT-2000 (first entry)

XX

DE Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

XX

OS Homo sapiens.

XX

OS Synthetic.

XX

PN WO200024782-A2.

XX

PD 04-MAY-2000.

XX

PF 25-OCT-1999; 99WO-US25044.

XX

PR 23-OCT-1998; 98US-0105371.

XX

PR 22-OCT-1999; 99US-0428082.

XX

XX (AMGE-) AMGEN INC.

PA

XX Feige U, Liu C, Cheetham J, Boone TC;

PI

XX WPI; 2000-350702/30.

DR

XX

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 PS Example 2; Page 327; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-Fl-(X2)b, where: Fl = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-Pl-(L2)d-P2, and L4 = are each
 CC -(L1)c-Pl-(L2)d-P2-(L3)e-P3, or -(L1)c-Pl-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where Pl, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 42 AA;
 Query Match 100.0%; Score 194; DB 21; Length 42;
 Best Local Similarity 100.0%; Pred. No. 2e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 36
 DB 7 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 42
 RESULT 8
 AAY96530
 ID AAY96530 standard; Protein; 42 AA.
 XX
 AC AAY96530;
 XX
 DT 04-SEP-2000 (first entry)
 DE Thrombopoietin mimetic peptide.
 XX
 KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TWP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.
 XX
 OS Synthetic.
 XX
 PN WO200024770-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24834.
 XX
 PR 23-OCT-1998; 98US-0105348.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Liu C, Feige U, Cheetham J;
 XX
 DR WPI; 2000-365108/31.
 DR N-PSDB; AAA29225.
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 PS Example 2A; Page 48; 91pp; English.

CC Overlapping oligonucleotides were used to construct a synthetic
 CC gene encoding a thrombopoietin mimetic peptide (TMP), which
 CC was then fused in-frame to the Fc region of the human IgG1 chain (see
 CC AAY96529). A compound which binds to an mpl receptor comprising a TMP
 CC dimer joined by a linker [TMP1-(L1)-TMP2], is new. TMP1 and TMP2
 CC are amino acid sequences varying from at least 10 to 14 residues in
 CC length comprising X2-X1-0, X2-X1-1, X2-X1-2, X2-X1-3, X2-X1-4,
 CC X1-X1-0, X1-X1-1, X1-X1-2, X1-X1-3, and X1-X1-4. X1 = I, A,
 CC V, L, S or R; X2 = E, D, K or V; X3 = G or A; X4 = P; X5 = T or S;
 CC X6 = L, I, V, A, or F; X7 = R or K; X8 = Q, N, or E; X9 = W, Y or F;
 CC X10 = L, I, V, A, F, M, or K; X11 = A, I, V, L, F, S, T, K, H, or E;
 CC X12 = A, I, V, L, F, G, S, or Q; X13 = R, K, T, V, N, Q or G; X14 =
 CC A, I, V, L, F, R, E, or G; L1 = linker comprising 1 to 20 amino
 CC acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl
 CC receptor which mediates the activity of endogenous thrombopoietin. The
 CC TWP are useful for increasing the production of platelets or platelet
 CC precursors (e.g. megakaryocytes) in a mammal, which is useful for
 CC treatment of diseases which involve thrombocytopenia, e.g. aplastic
 CC anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus
 CC associated ITP, and systemic lupus erythematosus.
 XX
 SQ Sequence 42 AA;
 Query Match 100.0%; Score 194; DB 21; Length 42;
 Best Local Similarity 100.0%; Pred. No. 2e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 36
 DB 7 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAAARA 42
 RESULT 9
 AAB17311
 ID AAB17311 standard; Peptide; 60 AA.
 XX
 AC AAB17311;
 XX
 DT 31-OCT-2000 (first entry)
 DE Synthetic TMP-TWP-Fc gene construction peptide SEQ ID NO:385.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 XX
 DR WPI; 2000-350702/30.
 XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 2; Page 331; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where F1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 60 AA;

Query Match 100.0%; Score 194; DB 21; Length 60;
 Best Local Similarity 100.0%; Pred. No. 2.8e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 ||||||||||||||||||||||||||||||||||||
 Db 2 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 37

RESULT 10

AAB16960
 ID AAB16960 standard; Protein; 269 AA.

AC AAB16960;

XX 31-OCT-2000 (first entry)

DE TMP-TMP-Fc protein sequence SEQ ID NO:10.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.
 OS Synthetic.

PN WO200024782-A2.

XX 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI: 2000-350702/30.

DR N-PSDB; AAA69446.

XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and

PT autoimmune diseases -

XX Example 2; Page 185-186; 608pp; English.

PS

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where F1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 269 AA;

Query Match 100.0%; Score 194; DB 21; Length 269;
 Best Local Similarity 100.0%; Pred. No. 1.3e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 ||||||||||||||||||||||||||||||||||||
 Db 2 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 37

RESULT 11

AAY96531
 ID AAY96531 standard; Protein; 269 AA.

AC AAY96531;

XX 04-SEP-2000 (first entry)

DE Human IgG1 Fc TMP fusion protein.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

OS Homo sapiens.

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

PR 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI: 2000-365108/31.

DR N-PSDB; AAA29229.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

PS Example 2A; Page 49-50; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_TMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
 CC X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and

CC X_1-X_1-4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
 CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
 CC or E; X_9 = W, Y or F; X_10 = L, I, V, A, F, M or K; X_11 = A, I, V,
 CC L, F, S, T, K, H, or E; X_12 = A, L, V, L, F, G, S, or Q; X_13 = R, K,
 CC T, V, N, Q or G; X_14 = A, I, V, L, F, T, R, E, or G; L_1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-wpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX SQ Sequence 269 AA;

Query Match 100.0%; Score 194; DB 21; Length 269;
 Best Local Similarity 100.0%; Pred. NO. 1.3e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAARA 36
 ||||||||||||||||||||||||||||||||||||
 Db 234 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAARA 269

RESULT 12

AAB16959
 ID AAB16959 standard; Protein: 268 AA.

XX AC AAB16959;

XX DT 31-OCT-2000 (first entry)

XX DE FC-TMP-TMP protein sequence SEQ ID NO:8.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI: 2000-350702/30.

XX DR N-PSDB; AAA69445.

XX PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX PS Example 2; Page 182-183; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present inventions are
 CC be used for producing pharmaceutical compositions. The compositions can
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX SQ Sequence 268 AA;

Query Match 97.9%; Score 190; DB 21; Length 268;
 Best Local Similarity 100.0%; Pred. NO. 3.8e-15;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 35
 ||||||||||||||||||||||||||||||||||||
 Db 234 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 268

RESULT 13

AAB17301

ID AAB17301 standard; Peptide: 36 AA.

XX AC AAB17301;

XX DT 31-OCT-2000 (first entry)

XX DE TPO-mimetic peptide sequence SEQ ID NO:357.

XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX DR WPI: 2000-350702/30.

XX PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX PS Example 1; Page 321; 608pp; English.

XX CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX
 SQ Sequence 36 AA;
 Query Match 95.4%; Score 185; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 2e-15; Indels 0; Gaps 0;
 Matches 35; Conservative 0; Mismatches 1;
 QY 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||

RESULT 16
 AAB17307
 ID AAB17307 standard; peptide; 36 AA.
 XX
 AC AAB17307;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:363.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.
 XX WO200024782-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 25-OCT-1999; 99WO-US25044.
 XX
 XX PR 23-OCT-1998; 98US-0105371.
 XX
 XX PR 22-OCT-1999; 99US-0428082.
 XX
 XX PA (AMGE-) AMGEN INC.
 XX
 XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX
 XX DR WPI; 2000-350702/30.
 XX
 XX PT Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX
 XX PS Example 1; Page 324; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX
 SQ Sequence 36 AA;
 Query Match 95.4%; Score 185; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 2e-15; Indels 0; Gaps 0;
 Matches 35; Conservative 0; Mismatches 1;
 QY 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 1 IEPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||

RESULT 17
 AAY96524
 ID AAY96524 standard; peptide; 36 AA.
 XX
 AC AAY96524;
 XX
 DT 04-SEP-2000 (first entry)
 XX
 DE Thrombopoietin mimetic peptide compound 5.
 XX
 KW Thrombopoietin; mimetic; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
 XX Synthetic.

XX Key Location/Qualifiers
 XX Modified-site 1 /note= "optionally linked to an Fc molecule"
 XX Peptide 1..14 /label= TMP_1
 XX Disulfide-bond 9..31 /note= "optional"
 XX Peptide 15..22 /label= linker
 XX Peptide 23..36 /label= TMP_2

XX WO200024770-A2.
 XX
 XX PD 04-MAY-2000.
 XX
 XX PF 22-OCT-1999; 99WO-US24834.
 XX
 XX PR 23-OCT-1998; 98US-0105348.
 XX
 XX PA (AMGE-) AMGEN INC.
 XX
 XX PI Liu C, Feige U, Cheetham J;
 XX
 XX DR WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 XX production of platelets or platelet precursors, useful for treatment of
 XX diseases which involve thrombocytopenia
 XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 XX mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L1)-TMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

CC 10 to 14 residues in length comprising X₂-X₁-X₀, X₂-X₁-X₁, X₂-X₁-X₁-X₂,
 CC X₂-X₁-X₁-X₃, X₂-X₁-X₁-X₄, X₁-X₁-X₁-X₁, X₁-X₁-X₁-X₂, X₁-X₁-X₁-X₃, and
 CC X₁-X₁-X₁-X₄. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K,
 CC T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;
 SQ Query Match 95.4%; Score 185; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 2e-15;
 Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAARA 36
 |||||
 Db 1 IEGPTLRQWLAAAGGGGGGIEGPTLRQWLAARA 36
 |||||

RESULT 18
 AAB17294
 ID AAB17294 standard; Peptide; 37 AA.
 XX AC AAB17294;
 XX DT 31-OCT-2000 (first entry)
 XX DE TPO-mimetic peptide sequence SEQ ID NO:350.
 XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 XX OS Synthetic.

XX WO200024782-A2.
 XX PN 04-MAY-2000.
 XX PD 25-OCT-1999; 99WO-US25044.
 XX PF 23-OCT-1998; 98US-0105371.
 XX PR 22-OCT-1999; 99US-0428082.
 XX PA (AMGE-) AMGEN INC.
 XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 37 AA;
 SQ Query Match 94.6%; Score 183.5; DB 21; Length 37;
 Best Local Similarity 97.3%; Pred. No. 3.1e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 IEGPTLRQWLAAARA-GGGGGGGGIEGPTLRQWLAARA 36
 |||||
 Db 1 IEGPTLRQWLAAARAGGGGGGGGIEGPTLRQWLAARA 37
 |||||

RESULT 19
 AAB17295
 ID AAB17295 standard; Peptide; 38 AA.
 XX AC AAB17295;
 XX DT 31-OCT-2000 (first entry)
 XX DE TPO-mimetic peptide sequence SEQ ID NO:351.
 XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.
 XX WO200024782-A2.
 XX PN 04-MAY-2000.
 XX PD 25-OCT-1999; 99WO-US25044.
 XX PF 23-OCT-1998; 98US-0105371.
 XX PR 22-OCT-1999; 99US-0428082.
 XX PA (AMGE-) AMGEN INC.
 XX PI Feige U, Liu C, Cheetham J, Boone TC;
 XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 XX pharmacologically active peptides, useful for treating cancer and
 XX autoimmune diseases -
 XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX SQ Sequence 38 AA;

Query Match 94.3%; Score 183; DB 21; Length 38;

Best Local Similarity 94.7%; Pred. No. 3.6e-15; Mismatches 0; Indels 2; Gaps 1;

QY 1 IEPTLRQWLAAARA--GGGGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 1 IEPTLRQWLAAARAAGGGGGGIEGPTLRQWLAAARA 38

RESULT 20

AAB17304

ID AAB17304 standard; Peptide; 39 AA.

XX AC AAB17304;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:360.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 323; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX SQ Sequence 39 AA;

Query Match 94.1%; Score 182.5; DB 21; Length 39;

Best Local Similarity 92.3%; Pred. No. 4.2e-15; Mismatches 0; Indels 3; Gaps 1;

QY 1 IEPTLRQWLAAARAGG--GGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 1 IEPTLRQWLAAARAGGKPEGGGGIEGPTLRQWLAAARA 39

RESULT 21

AAB17305

ID AAB17305 standard; Peptide; 39 AA.

XX AC AAB17305;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:361.

Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 323; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC sequences used in the exemplification of the present invention.
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX
 XX
 SQ Sequence 39 AA;
 Query Match 94.1%; Score 182.5; DB 21; Length 39;
 Best Local Similarity 92.3%; Pred. No. 4.2e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 IEGPTLRQWLAAARAGG---GGGGGEGPTLRQWLAAARA 36
 |||||
 DB 1 IEGPTLRQWLAAARAGGCGPEGGGGEGPTLRQWLAAARA 39

RESULT 22
 AAB17306
 ID AAB17306 standard; Peptide; 36 AA.
 XX
 AC AAB17306;

XX
 DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:362.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases.

XX Example 1; Page 324; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 93.8%; Score 182; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 4.5e-15;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGCGGGGEGPTLRQWLAAARA 36
 |||||
 DB 1 IEGPTLRQWLAAARAGGCGGGGEGPTLRQWLAAARA 36

RESULT 23

AA96526

ID AAY96526 standard; peptide; 36 AA.

XX

XX AAY96526;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 7.

XX Thrombopoietin; mimetic; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14

FT Peptide /label= TMP_1

FT Peptide 15..18

FT Peptide /label= linker

FT Peptide 19..32

FT Peptide /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
 CC X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
 CC X_1-X_1_4. X_1 = L, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
 CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
 CC or E; X_9 = W, Y or G; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
 CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
 CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and

CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

Query Match 93.8%; Score 182; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 4.5e-15;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTTLROWLAARAGGGGGGIEPTTLROWLAARA 36
 |||||
 Db 1 IEPTTLROWLAARAGGGGGGIEPTTLROWLAARA 36

RESULT 24

AAB17296
 ID AAB17296 standard; Peptide: 42 AA.

XX AC AAB17296;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:352.

XX Modified peptide: therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independent
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 42 AA;

Query Match 93.3%; Score 181; DB 21; Length 42;
 Best Local Similarity 85.7%; Pred. No. 6.9e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 1 IEPTTLROWLAARA-----GGGGGGGIEPTTLROWLAARA 36
 |||||
 Db 1 IEPTTLROWLAARAGGGGGGIEPTTLROWLAARA 42

RESULT 25

AAB17292

ID AAB17292 standard; Peptide: 35 AA.

XX AC AAB17292;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:348.

XX Modified peptide: therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 317-318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independent
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 35 AA;
Query Match 91.5%; Score 177.5; DB 21; Length 35;
Best Local Similarity 97.2%; Pred. No. 1.5e-14;
Matches 35; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
|||||
Db 1 IEGPTLRQWLAAARA-GGGGGGIEGPTLRQWLAAARA 35

RESULT 26
AAB17302
ID AAB17302 standard; Peptide; 40 AA.
AC AAB17302;
XX
XX 31-OCT-2000 (first entry)
XX
XX TPO-mimetic peptide sequence SEQ ID NO:358.
DE
DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
XX WO200024782-A2.
XX
XX 04-MAY-2000.
XX
XX 25-OCT-1999; 99WO-US25044.
XX
XX 23-OCT-1998; 98US-0105371.
XX
XX 22-OCT-1999; 99US-0428082.
XX
XX (AMGE-) AMGEN INC.
XX
XX Feige U, Liu C, Cheetham J, Boone TC;
XX
XX WPI; 2000-350702/30.
XX
XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -

Example 1; Page 322; 608pp; English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

XX
SQ Sequence 40 AA;
Query Match 89.7%; Score 174; DB 21; Length 40;
Best Local Similarity 87.5%; Pred. No. 4.5e-14;
Matches 35; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
|||||
Db 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 40

RESULT 27
AAB17291
ID AAB17291 standard; Peptide; 34 AA.
XX
XX AAB17291;
XX
XX 31-OCT-2000 (first entry)
XX
XX TPO-mimetic peptide sequence SEQ ID NO:347.
DE
DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
XX WO200024782-A2.
XX
XX 04-MAY-2000.
XX
XX 25-OCT-1999; 99WO-US25044.
XX
XX 23-OCT-1998; 98US-0105371.
XX
XX 22-OCT-1999; 99US-0428082.
XX
XX (AMGE-) AMGEN INC.
XX
XX Feige U, Liu C, Cheetham J, Boone TC;
XX
XX WPI; 2000-350702/30.
XX
XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -

Example 1; Page 317; 608pp; English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Query Match 88.1%; Score 171; DB 21; Length 34;
 Best Local Similarity 94.4%; Pred. No. 8.6e-14;
 Matches 34; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 1 IEGPTLRQWLAAARA--GGGGGIEGPTLRQWLAAARA 34

RESULT 28

AAB17298
 ID AAB17298 standard; Peptide; 36 AA.

XX AC AAB17298;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:354.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 320; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 86.6%; Score 168; DB 21; Length 36;

Best Local Similarity 94.4%; Pred. No. 2.1e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
 |||||
 Db 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 29

AAB17299
 ID AAB17299 standard; Peptide; 36 AA.

XX AC AAB17299;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:355.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX OS Synthetic.

XX PN WO200024782-A2.

XX PD 04-MAY-2000.

XX PF 25-OCT-1999; 99WO-US25044.

XX PR 23-OCT-1998; 98US-0105371.

XX PR 22-OCT-1999; 99US-0428082.

XX PA (AMGE-) AMGEN INC.

XX PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 320-321; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 86.6%; Score 168; DB 21; Length 36;

Best Local Similarity 94.4%; Pred. No. 2.1e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 IEGPTLRQWLARAGGGGGGGIEGPTLRQWLARA 36 QY
||||| ||||||| ||||||| ||||||| |||||||
1 IEGPTLRQWLARAGGGGGGGIEGPTLRQWLARA 36 Db
||||| ||||||| ||||||| ||||||| |||||||

RESULT 30

RECOLL 30
AAY96521
ID AAY96521 standard: peptide: 36 AA.

AA96521:

04-SEP-2000 (first entry)

DE Cyclic or linear thrombopoietin mimetic peptide compound 2.

Thrombopoietin; mimetic; TMP; platelet; megakaryocyte; production;
KW
anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW
immunosuppressive; anti-inflammatory; linker; cyclic; linear.
KW

OS Synthetic.

Key	Location/Qualifiers
Modified-site	1
FT	/note= "optionally linked to an Fc molecule"
FT	1..14
Peptide	/label= TMP_1
FT	9..31
Disulfide-bond	/note= "optional"
FT	15..22
Peptide	/label= linker
FT	23..36
Peptide	/label= TMP_2
FT	

PN WO200024770-A2.

04-MAY-2000.

PF 22-OCT-1999; 99WO-US24834.

PR 23-OCT-1998; 98US-0105348.

PA (AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;

DR WPI; 2000-365108/31.

Thrombopoietic peptides which activate mpl receptors and increase the production of platelets or platelet precursors, useful for treatment of diseases which involve thrombocytopenia

PS Claim 16; page 61; 91pp; English.

A compound which binds to an mpl receptor comprising a thrombopoietin mimetic peptide (TMP) dimer joined by a linker [TMP₁-(L₁)₁-nTMP₂], is new. TMP₁ and TMP₂ are amino acid sequences varying from at least 12 to 14 residues in length comprising X₂-X₁-L₀, X₂-X₁-L₁, X₂-X₁-L₂, X₂-X₁-L₃, X₂-X₁-L₄, X₁-X₁-L₀, X₁-X₁-L₁, X₁-X₁-L₂, X₁-X₁-L₃, and X₁-X₁-L₄. X₁=I, A, V, L, S or R; X₂=E, D, K or V; X₃=G or A; X₄=L or F; X₅=T or S; X₆=L, I, V, A or F; X₇=R or K; X₈=O, N, or E; X₉=W, Y or F; X₁₀=L, I, V, A, F, M, or K; X₁₁=A, I, V, L, F, S, T, Q, R, H, or E; X₁₂=A, I, V, L, F, G, S, or Q; X₁₃=R, K, T, V, N, Q or G; X₁₄=A, I, V, L, F, T, R, E, or G; L₁=linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and activate the c-mpl receptor which mediates the activity of endogenous thrombopoietin. The TMPs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

AA	Sequence	36 AA;
SQ		

Query Match 86.6%; Score 168; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 2.le-13;
Matches 34; Conservative 0; Mismatches 2; Indels

Qy 1 IEGPTLRQLAARAGGGGGGIEGPTLRQWLAARA 36
||||||| ||||||| ||||||| ||||||| |||||||
Dδ 1 IEGPTLROCLAARAGGGGGGIEGPTLROCILAARA 36

Search completed: October 9, 2002, 08:58:55
Job time : 16.1874 secs


```
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /product= "Ava"
; US-08-764-640-231

Query Match 39.4%; Score 76.5; DB 2; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0058;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 2 EGPTRLQWLARAGGGGGGIEGPTLRQWLA 33
; :|||||: | :|||||:
Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 2
US-09-244-298A-231
; Sequence 231, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: 514
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /product= "Ava"
; US-09-244-298A-231

Query Match 39.4%; Score 76.5; DB 3; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0058;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 2 EGPTRLQWLARAGGGGGGIEGPTLRQWLA 33
; :|||||: | :|||||:
Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 3
US-09-516-704-231
; Sequence 231, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /product= "Ava"
; US-09-516-704-231

Query Match 39.4%; Score 76.5; DB 4; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0058;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

QY 2 EGPTRLQWLARAGGGGGGIEGPTLRQWLA 33
; :|||||: | :|||||:
Db 2 DGPTLREWISFXA-----DGPTLREWIS 24

RESULT 4
US-08-764-640-13
; Sequence 13, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
```



```
;
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESS: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-13

Query Match 37.6%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy- 1 IEGPTLRQWLAARA 14
    |||||
Db  1 IEGPTLRQWLAARA 14

RESULT 5
US-08-764-640-193
; Sequence 193, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprience, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
```

```
;
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESS: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-193

Query Match 37.6%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 IEGPTLRQWLAARA 14
    |||||
Db  1 IEGPTLRQWLAARA 14

RESULT 6
US-08-973-225-13
; Sequence 13, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirila, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haselden, Sherril S.
; APPLICANT: Mattheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESS: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
```

;; FILING DATE: 04-Dec-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hrubiec, Robert T.
;; REGISTRATION NUMBER: 36,392
;; REFERENCE/DOCKET NUMBER: PK3065USW
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919-248-1000
;; INFORMATION FOR SEQ ID NO: 13:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14
|||||

RESULT 7
US-08-973-225-193
; Sequence 193, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Matheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-Dec-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-08-973-225-193

Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14
|||||

RESULT 8
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depreince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAAARA 14
Db 1 IEGPTLRQWLAAARA 14
|||||

RESULT 9
US-09-244-298A-193
; Sequence 193, Application US/09244298A

Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-244-298A-193

Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
Db 1 IEPTLRQWLAARA 14

RESULT 10
US-09-516-704-13
; Sequence 13, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13

Query Match 37.6%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
Db 1 IEPTLRQWLAARA 14

RESULT 11
US-09-516-704-193
; Sequence 193, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match 37.6%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 12
US-08-764-640-17
; Sequence 17, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depreince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-185

Query Match 37.6%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15
```

```
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-17

Query Match 37.6%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 13
US-08-764-640-185
; Sequence 185, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Depreince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-185

Query Match 37.6%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15
```

RESULT 14

US-08-973-225-17
; Sequence 17, Application US/08973225A

; Patent No. 6083913

; GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirila, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973,225A

FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 17:

US-08-973-225-17

Query Match

Best Local Similarity 37.6%; Score 73; DB 3; Length 15;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLROWLAARA 14

Db 1 IEGPTLROWLAARA 14

RESULT 15

US-08-973-225-185

; Sequence 185, Application US/08973225A

; Patent No. 6083913

; GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirila, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.

Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973,225A

FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-08-973-225-185

Query Match

Best Local Similarity 37.6%; Score 73; DB 3; Length 15;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLROWLAARA 14

Db 2 IEGPTLROWLAARA 15

RESULT 16

US-09-244-298A-17

; Sequence 17, Application US/09244298A

; Patent No. 6121238

; GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17

Query Match 37.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | | | |
DB 1 IEGPTLRQWLAARA 14

RESULT 17
US-09-244-298A-185
Sequence 185, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-185

Query Match 37.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | | | |
DB 2 IEGPTLRQWLAARA 15

RESULT 18
US-09-516-704-17
Sequence 17, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match 37.6%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPTLRQWLAARA 14
| | | | | | | | | | | | | | | | | |

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Glaxo Wellcome
;; STREET: Five Moore Drive, P.O. Box 13398
;; CITY: Research Triangle Park
;; STATE: NC
;; COUNTRY: USA
;; ZIP: 27709
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/764,640
;; FILING DATE: 11-DEC-1996
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hrubiec, Robert T.
;; REGISTRATION NUMBER: 36,392
;; REFERENCE/DOCKET NUMBER: PK3281
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919-248-1000
;; INFORMATION FOR SEQ ID NO: 194:
;; LENGTH: 16 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-764-640-194

Query Match 37.6%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

RESULT 22
US-08-764-640-232
;; Sequence 232, Application US/08764640
;; Patent No. 5869451
;; Patent No. 5869451 5837683
;; GENERAL INFORMATION:
;; APPLICANT: Dower, William J.
;; APPLICANT: Barrett, Ronald W.
;; APPLICANT: Cwirla, Steven E.
;; APPLICANT: Gates, Christian
;; APPLICANT: Schatz, Peter J.
;; APPLICANT: Balasubramanian, Palaniappan
;; APPLICANT: Wagstrom, Christopher R.
;; APPLICANT: Hendren, Richard W.
;; APPLICANT: Depnence, Randolph B.
;; APPLICANT: Podduturi, Surekha
;; APPLICANT: Yin, Qun
;; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;; NUMBER OF SEQUENCES: 244
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Glaxo Wellcome
;; STREET: Five Moore Drive, P.O. Box 13398
;; CITY: Research Triangle Park
;; STATE: NC
;; COUNTRY: USA
;; ZIP: 27709
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/764,640
;; FILING DATE: 11-DEC-1996
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hrubiec, Robert T.
;; REGISTRATION NUMBER: 36,392
;; REFERENCE/DOCKET NUMBER: PK3281
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919-248-1000
;; INFORMATION FOR SEQ ID NO: 232:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 16 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-764-640-232

Query Match 37.6%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

RESULT 23
US-08-973-225-18
;; Sequence 18, Application US/08973225A
;; Patent No. 6083913
;; GENERAL INFORMATION:
;; APPLICANT: Dower, William J.
;; APPLICANT: Barrett, Ronald W.
;; APPLICANT: Cwirla, Steven E.
;; APPLICANT: Duffin, David J.
;; APPLICANT: Gates, Christian
;; APPLICANT: Haselden, Sherrill S.
;; APPLICANT: Mattheakis, Larry C.
;; APPLICANT: Schatz, Peter J.
;; APPLICANT: Wagstrom, Christopher R.
;; APPLICANT: Wrighton, Nicholas C.
;; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
;; NUMBER OF SEQUENCES: 232
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Glaxo Wellcome
;; STREET: Five Moore Drive, P.O. Box 13398
;; CITY: Research Triangle Park
;; STATE: NC
;; COUNTRY: USA
;; ZIP: 27709
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/973,225A
;; FILING DATE: 04-DEC-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hrubiec, Robert T.
;; REGISTRATION NUMBER: 36,392
;; REFERENCE/DOCKET NUMBER: PK3065USW
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 919-248-1000
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 16 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide

FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLROWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLROWLAARA 14

RESULT 24
US-08-973-225-194
Sequence 194, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLROWLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPTLROWLAARA 15

RESULT 25

US-08-973-225-220
Sequence 220, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 220:
SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLROWLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPTLROWLAARA 15

RESULT 26

US-09-244-298A-18
Sequence 18, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.

APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
US-09-244-298A-18

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQWLAARA 14

RESULT 27
US-09-244-298A-194
Sequence 194, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQWLAARA 15

RESULT 28
US-09-244-298A-232
Sequence 232, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-232

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

RESULT 29

US-09-516-704-18
Sequence 18, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product= "Beta-ala"

SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-516-704-18

Query Match 37.6%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
Db 1 IEGPTLRQWLAARA 14

RESULT 30

US-09-516-704-194

Sequence 194, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 194:

US-09-516-704-194

Query Match 37.6%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGPTLRQWLAARA 14
Db 2 IEGPTLRQWLAARA 15

Search completed: October 9, 2002, 09:06:30
Job time : 6.98595 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 8.09368 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838C-24

Perfect score: 194

Sequence: 1 IEPTLRQLAARAGGGGGGIEGPTLRQLAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	35.6	500	2 T20961	hypothetical prote
2	68.5	35.3	488	2 G87033	probable ATP/Gnp-b
3	68.5	35.3	518	2 S72938	hflx protein - Myc
4	66.5	34.3	495	2 D70505	probable Hflx - My
5	64	33.0	201	2 T49792	hypothetical prote
6	64	33.0	331	2 T26807	hypothetical prote
7	64	33.0	333	2 T26808	hypothetical prote
8	63.5	32.7	619	1 KSNCLT	laccase (EC 1.10.3
9	63.5	32.7	619	1 KSNCLT	hypothetical 20.2K
10	63	32.5	201	2 T01094	phosphatidylinosit
11	63	32.5	490	2 T09084	neurotrophin-4 pre
12	62.5	32.2	209	2 E42687	subtilisin-like pr
13	61.5	31.7	487	2 B39490	subtilisin-like pr
14	61.5	31.7	652	1 JC2191	subtilisin-like pr
15	61.5	31.7	962	2 JC5571	subtilisin-like pr
16	61.5	31.7	969	1 A39490	subtilisin-like pr
17	61.5	31.7	975	2 JC5570	subtilisin-like pr
18	61	31.4	415	2 D96664	hypothetical prote
19	61	31.4	443	1 S29334	transcription fact
20	61	31.4	445	1 S31224	transcription fact
21	61	31.4	593	1 KRH00	keratin 10, type I
22	61	31.4	777	2 S65543	3',5'-cyclic-nucle
23	61	31.4	1168	1 MWAXIC	myosin heavy chain
24	60.5	31.2	210	2 A42687	neurotrophin-4 pre
25	60.5	31.2	864	2 A48266	protein-tyrosine k
26	60	30.9	285	2 S69312	probable membrane
27	60	30.9	323	2 S20099	transforming prote
28	60	30.9	569	1 KRMSE1	keratin, 59K type
29	60	30.9	649	2 S58064	hdc protein - frul

30	60	30.9	806	2 T13690	hypothetical prote
31	60	30.9	888	2 T58378	tyrosine kinase -
32	60	30.9	962	2 T04124	receptor-like prot
33	60	30.9	1325	2 T13386	hypothetical prote
34	59.5	30.7	327	2 B84781	hypothetical prote
35	59.5	30.7	339	2 T06612	hypothetical prote
36	59.5	30.7	403	2 A53662	homeotic protein H
37	59.5	30.7	443	2 E96495	hypothetical prote
38	59.5	30.7	867	2 S57795	probable deoxyribo
39	59	30.4	80	2 T10550	hypothetical prote
40	59	30.4	199	2 T48099	hypothetical prote
41	59	30.4	250	2 H85067	hypothetical prote
42	59	30.4	270	2 T35365	hypothetical prote
43	59	30.4	346	1 S35500	heterogeneous ribo
44	59	30.4	367	2 JC6087	helix-loop-helix t
45	59	30.4	396	2 T49109	glycine-rich prote

ALIGNMENTS

RESULT 1

T20961

hypothetical protein F15B9.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T20961

R:Percy, C.

Submitted to the EMBL Data Library, August 1996

A:Reference number: Z19351

A:Accession: T20961

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-500 <WIL>

A:Cross-references: EMBL:Z78013; PIDN:CAB01420.1; GSPDB:GN00023; CESP:F15B9.5

A:Experimental source: clone F15B9

C:Genetics:

A:Gene: CESP:F15B9.5

A:Map position: 5

A:Introns: 46/3; 63/3; 125/2; 162/2; 283/3; 391/1; 446/1

Query Match

Best Local Similarity 35.6%; Score 69; DB 2; Length 500;

Matches 13; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 GPTLRQLAARAGGGGGGIEG 25

Db 429 GSNLGRFLSNRGGGGGGG 451

RESULT 2

G87033

probable ATP/GTP-binding protein [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: G87033

R:Coile, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;

A>Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: G87033

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-488 <STO>

A:Cross-references: GB:AL450380; NID:g13093026; PIDN:CAC31378.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML0997

C:Superfamily: GTP-binding protein hflx; translation elongation factor Tu homology

Query Match

Score 35.3%; Score 68.5; DB 2; Length 488;

Best Local Similarity 46.7%; Pred. No. 3.7;
Matches 14; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGGGGGIEGP 26
| | | | : | | | | | | | | | |
Db 189 PRLRGESMSRQVGRAGSGGGVGLRGP 218

RESULT 3

S72938
hflX protein - Mycobacterium leprae
N:Alternate names: B2235_C2_202 protein
C:Species: Mycobacterium leprae
C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C:Accession: S72938
R:Smith, D.R.; Robinson, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid B2235.
A:Reference number: S72587
A:Accession: S72938
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-518 <SMI>
A:Cross-references: EMBL:U00019; NID:g467079; PIDN:AAA17274.1; PID:g467091
C:Genetics:
A:Start codon: GTG
C:Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match 35.3%; Score 68.5; DB 2; Length 518;
Best Local Similarity 46.7%; Pred. No. 3.9;
Matches 14; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGGGGGIEGP 26
| | | | : | | | | | | | | | |
Db 219 PRLRGESMSRQVGRAGSGGGVGLRGP 248

RESULT 4

D70505
Probable HflX - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 02-Sep-2000
C:Accession: D70505
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Hollroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70505
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-495 <COL>
A:Cross-references: GB:Z98209; GB:AL123456; NID:g3261838; PIDN:CAB10901.1; PID:e332282;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: hflX
C:Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match 34.3%; Score 66.5; DB 2; Length 495;
Best Local Similarity 46.7%; Pred. No. 6;
Matches 14; Conservative 1; Mismatches 8; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGGGGGIEGP 26
| | | | : | | | | | | | | | |
Db 199 PRLRGESMSRQVGRAGSGGGVGLRGP 228

RESULT 5

T49792
hypothetical protein B9J10.290 [imported] - Neurospora crassa
C:Species: Neurospora crassa

C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49792
R:Schulte, U.; Align, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, May 2000
A:Reference number: T25022
A:Accession: T49792
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <SCH>
A:Cross-references: EMBL:ALJ56324; GSPDB:GNO0116; NCSP:B9J10.290
A:Experimental source: BAC clone B9J10; strain OR74A
C:Genetics:
A:Gene: NCSP:B9J10.290
A:Map position: 6

Query Match 33.0%; Score 64; DB 2; Length 201;
Best Local Similarity 57.1%; Pred. No. 4.9;
Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 13 RAGGGGGGGGIEGPTLRQWLA 33
| | | | | | | | | | | | | | | |
Db 74 RGGGGGGGGVNG-----RWSA 90

RESULT 6

T26807
hypothetical protein Y41C4A.4a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
C:Accession: T26807
R:Steward, C.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20269
A:Accession: T26807
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-331 <WIL>
A:Cross-references: EMBL:AL032627; PIDN:CAB54381.1; CESP:Y41C4A.4a
A:Experimental source: clone Y41C4A
C:Genetics:
A:Gene: CESP:Y41C4A.4a
A:Introns: 24/3; 50/2; 81/3; 159/1; 228/1; 292/3
C:Superfamily: fos/jun DNA-binding domain homology

Query Match 33.0%; Score 64; DB 2; Length 331;
Best Local Similarity 76.9%; Pred. No. 7.7;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27
| | | | | | | | | | | | | | | |
Db 167 GGGGGGGGVPGPS 179

RESULT 7

T26808
hypothetical protein Y41C4A.4b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
C:Accession: T26808
R:Steward, C.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20269
A:Accession: T26808
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-333 <WIL>
A:Cross-references: EMBL:AL032627; PIDN:CAB54382.1; CESP:Y41C4A.4b
A:Experimental source: clone Y41C4A
C:Genetics:
A:Gene: CESP:Y41C4A.4b
A:Introns: 24/3; 50/2; 81/3; 161/1; 230/1; 294/3
C:Superfamily: fos/jun DNA-binding domain homology

C:Genetics: 86/3
A:Introns: 86/3
C:Superfamily: laccase
C:Keywords: copper; glycoprotein; oxidoreductase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-49/Domain: propeptide #status predicted <PRO>
F:50-619/Product: laccase #status predicted <MAT>
F:84-215/Domain: amino-terminal beta-barrel #status predicted <BBI>
F:216-372/Domain: middle beta-barrel #status predicted <BB2>
F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F:139,282,295,340,422,444/Binding site: carbonylde (Asn) (covalent) #status predicted
F:144,480/Binding site: copper (His) (type 2) #status predicted
F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status predicted
F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 32.7%; Score 63.5; DB 1; Length 619;
Best Local Similarity 57.7%; Pred. No.15;
Matches 15; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 7 ROWLAARAGGGGGGIEGPTLRQ-W 31
||| ||||| |||||
Ddb 39 RDSQAERYGGGGGCGNSPTNRQW 64

RESULT 10
JQ1094
hypotheical 20.2K protein - tomato ringspot virus
C:Species: tomato ringspot virus
C:Species: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 08-Oct-1999
C:Accession: JQ1094
A:Reference number: JQ1093; MUID:91311402
A:Accession: JQ1094
A:Status: translation not shown
A:Molecule type: genomic RNA
A:Residues: 1-201 <ROT>
A:Cross-references: GB:D12477; GB:D12477; PIDN:g222674; PIDN:BAA02044.1; PID:d1002526;
A:Experimental source: strain raspberry

Query Match 32.5%; Score 63; DB 2; Length 201;
Best Local Similarity 61.5%; Pred. No. 6.2;
Matches 16; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 13 RAGGGGGGGIE----GPTLRQWLAA 34
||||||| |||||
Ddb 13 RAGGGGGGGGKVFVFRAGRTLLKVLKA 38

RESULT 11
T09084
phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)
C:Species: Chlamydomonas reinhardtii
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09084
A:Reference number: Z16411; MUID:98281574
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-490 <MOL>
A:Cross-references: EMBL:U97663; NID:g2109290; PIDN:AAG50018.1; PID:g2109291
A:Experimental source: strain cw-15
A:Genetics:
A:Introns: 265/3; 331/3; 370/3; 455/1; 481/3

Query Match 32.5%; Score 63; DB 2; Length 490;
Best Local Similarity 48.6%; Pred. No. 14;
Matches 17; Conservative 2; Mismatches 6; Indels 10; Gaps 3;

Nucleic Acids Res. 21, 253-258, 1993
 A:Title: cDNA cloning of human N-Oct 3, a nervous-system specific POU domain transcript
 A:Reference number: S30296; MUID:93181199
 A:Accession: S30296

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-25,'G',27-443 <SCW>

A:Cross-references: EMBL:Z11933

A:Experimental source: tissue-type brain

C:Genetics:

A:Gene: GDB:POU3F2; OMF7

A:Cross-references: GDB:222816; OMIM:600494

A:Map position: 6q16-q16

C:Superfamily: transitive initiation factor Brn-1; homeobox homology; POU domain homology
 C:Keywords: alternative initiators; DNA binding; homeobox; nucleus; transcription regula

F:1-443/Product: transcription factor Brn-2 #status experimental <MAT1>

F:125-149/Region: glutamine-rich

F:151-165/Region: histidine/proline-rich

F:181-443/Product: transcription factor Oct-5a #status experimental <MAT2>

F:200-443/Product: transcription factor Oct-5b #status experimental <MAT3>

F:211-259/Region: histidine/proline-rich

F:269-336/Domain: POU domain homology <POU>

F:355-411/Domain: homeobox homology <HOX>

Query Match

Best Local Similarity 31.4%; Score 61; DB 1; Length 443;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 QWLAARAGGGGGGG 22

DB 60 QWITALSHGGGGGG 74

RESULT 20

S31224

transcription factor Brn-2 - mouse

N:Alternate names: class III POU domain protein brain-2

C:Species: Mus musculus (house mouse)

C:Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 22-Jun-1999

C:Accession: S31224

R:Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.

Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992

A:Title: Structure and evolution of four POU domain genes expressed in mouse brain.

A:Reference number: S31223; MUID:92228768

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-445 <HAR>

A:Cross-references: EMBL:M88300; NID:g200446; PIDN:AAA39961.1; PID:g200447

C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:68-90/Region: glycine-rich

F:125-151/Region: glutamine-rich

F:153-165/Region: histidine/proline-rich

F:213-261/Region: histidine/proline-rich

F:271-338/Domain: POU domain homology <POU>

F:357-413/Domain: homeobox homology <HOX>

Query Match

Best Local Similarity 31.4%; Score 61; DB 1; Length 445;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 QWLAARAGGGGGGG 22

DB 60 QWITALSHGGGGGG 74

RESULT 21

KRHUO

keratin 10, type I, cytoskeletal - human

N:Alternate names: cytokeratin 10

C:Species: Homo sapiens (man)

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 10-Dec-1999
 C:Accession: S02158; C38182; B38182; PC1102; S14666; S14669
 R:Rieger, M.; Franke, W.W.

J. Mol. Biol. 204, 841-856, 1988

A:Title: Identification of an orthologous mammalian cyokeratin gene. High degree of

A:Reference number: S02158; MUID:89125611

A:Molecule type: DNA

A:Residues: 1-593 <RIE>

A:Cross-references: EMBL:X14487; NID:928316; PIDN:CAA32649.1; PID:g28317

A:Experimental source: clone lambda-KH10-5

R:Korge, B.P.; Gan, S.Q.; McBride, O.W.; Mischke, D.; Steinert, P.M.

Proc. Natl. Acad. Sci. U.S.A. 89, 910-914, 1992

A:Title: Extensive size polymorphism of the human keratin 10 chain resides in the C-t

A:Reference number: A38182; MUID:92141228

A:Accession: C38182

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 452-593 <KOR1>

A:Cross-references: PIDN:AA21315.1; PID:g244509

A:Note: sequence extracted from NCBI backbone (NCBIP:79427)

A:Accession: B38182

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 452-463,'P',465-507,'Y',523-593 <KOR2>

A:Cross-references: PIDN:AA21314.1; PID:g244508

A:Note: sequence extracted from NCBI backbone (NCBIP:79431)

R:Kachenko, A.V.; Buchman, V.L.; Bliskovsky, V.V.; Shvets, Y.P.; Kisselev, L.L.

Gene 116, 245-251, 1992

A:Title: Exons I and VII of the gene (Ker10) encoding human keratin 10 undergo struct

A:Reference number: PC1102; MUID:92339897

A:Accession: PC1102

A:Molecule type: mRNA

A:Residues: 'G',198-407,'Q',409-450,'G',452-486,491-524,534-593 <TKA>

A:Cross-references: GB:M77663; NID:gl86628; PIDN:AAA59199.1; PID:gl86629

A:Experimental source: embryonic skin, clone HK51

R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.

Mol. Biol. Rep. 12, 277-283, 1987

A:Title: Sequence of a cDNA encoding human keratin No 10 selected according to struct

A:Reference number: S14666; MUID:88122104

A:Accession: S14666

A:Molecule type: mRNA

A:Residues: 130-278,'YV',281-311,'I',313-339,'V',341-373,'R',375-407,'Q',409-459,'RS'

56-579,'P',581-593 <DAR1>

A:Cross-references: EMBL:M19156; NID:gl86769

A:Note: the sequence from Fig. 3 is inconsistent with the nucleotide sequence from Fi

R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.

submitted to the EMBL Data Library, May 1988

A:Reference number: S14667

A:Accession: S14667

A:Molecule type: mRNA

A:Residues: 130-278,'YV',281-311,'I',313-339,'V',341-373,'R',375-407,'Q',409-459,'RS'

56-593 <DAR2>

A:Cross-references: EMBL:M19156; NID:gl86769; PIDN:AAA59468.1; PID:g307086

A:Note: the translated sequence in GenBank entry HUMKRT10A, release 111.0, differs fr

C:Genetics:

A:Gene: GDB:KRT10; KPP

A:Cross-references: GDB:118828; OMIM:148080

A:Map position: 17q12-17q21

A:Introns: 209/3; 237/2; 289/3; 343/3; 385/3; 458/2; 592/3

A:Note: this gene encodes variants with considerable length polymorphism

C:Complex: heterotetramer of two type I and two type II proteins, usually keratin 1 (

C:Superfamily: cytoskeletal keratin

C:Keywords: coiled coil; heterotetramer; intermediate filament; polymorphism

F:1-145/Domain: head <HEA>

F:146-456/Domain: helical rod #status predicted <ROD>

F:457-593/Domain: tail <TAI>

Query Match

Best Local Similarity 31.4%; Score 61; DB 1; Length 593;

Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

F:1-143/Domain: head <HED>
 F:1-143/Region: E1 and V1 subdomains
 F:144-457/Domain: rod <ROD>
 F:144-178/Region: coil 1A
 F:179-192/Region: linker 1
 F:193-293/Region: coil 1B
 F:294-309/Region: linker 12
 F:310-328/Region: coil 2A
 F:329-336/Region: linker 2
 F:337-457/Region: coil 2B
 F:395/Region: stutttter
 F:458-569/Domain: tail <END>
 F:458-569/Region: V2 and E2 subdomains

Query Match 30.9%; Score 60; DB 1; Length 569;

Best Local Similarity 43.5%; Pred. No. 33;

Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 7 ROWLAARAGGGGGGIEGPTLR 29

Db 9 KQFSSRSRGGGGGVRVSSPR 31

RESULT 29

S58064

hdc protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 24-Sep-1998

C:Accession: S58064

R:Weaver, T.A.; White, R.A.

submitted to the EMBL Data Library, July 1995

A:Description: hdc, an imaginal specific gene required for adult morphogenesis in Drosophila

A:Reference number: S58064

A:Accession: S58064

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-649 <WEA>

A:Cross-references: EMBL:Z50097; NID:g902623; PID:g902624

C:Genetics:

A:Gene: FlyBase:hdc

A:Cross-references: FlyBase:FBgn0010113

Query Match 30.9%; Score 60; DB 2; Length 649;

Best Local Similarity 76.9%; Pred. No. 37;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27

Db 220 GGGGGGGVNGNT 232

RESULT 30

TI13690

hypothetical protein EG0003.2 - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: TI13690

R:Murphy, L.; Harris, D.; Barrell, B.

submitted to the EMBL Data Library, November 1998

A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A:Reference number: 217699

A:Accession: TI13690

A:Status: preliminary; translated from GH/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-806 <MUR>

A:Cross-references: EMBL:AL031863; NID:el1331652; PID:el1355938; PIDN:CAA21318.1

A:Cross-references: FlyBase:FBgn0025833

A:Introns: 37/3; 448/3; 611/2; 690/3

A:Note: EG:EG0003.2

Query Match

Best Local Similarity 30.9%; Score 60; DB 2; Length 806;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPTLRQWLAARA 36

Db 100 GGGGGGGGPGGASITQAIQAA 121

Search completed: October 9, 2002, 09:05:03

Job time : 11.0937 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29577 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838c-24

Perfect score: 194

Sequence: 1 IEPTLRQLAARAGGGGGIEGPTLRQLAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63.5	32.7	619	1 LAC1_NEUCR	P06811 neurospora
2	63.5	32.7	619	1 LAC2_NEUCR	P10574 neurospora
3	63	32.5	201	1 YR21_TRSVR	P25245 tomato ring
4	62.5	32.2	209	1 NT5_RAT	P34131 rattus norv
5	62.5	32.2	266	1 SCO2_HUMAN	O43819 homo sapien
6	61.5	31.7	969	1 PAC4_HUMAN	P29122 homo sapien
7	61	31.4	394	1 FXD3_CHICK	P79772 gallus gall
8	61	31.4	443	1 OC3N_HUMAN	P20265 homo sapien
9	61	31.4	445	1 OC3N_MOUSE	P31360 mus musculu
10	61	31.4	584	1 CNAL_DROME	P12252 drosophila
11	61	31.4	593	1 K1CJ_HUMAN	P13645 homo sapien
12	61	31.4	1168	1 MYSC_ACACA	P10569 acanthamoeb
13	61	31.4	1178	1 PHYB_SORBI	P93527 sorghum bic
14	60.5	31.2	210	1 NT5_HUMAN	P34130 homo sapien
15	60.5	31.2	864	1 KLFK_HUMAN	P29376 homo sapien
16	60	30.9	323	1 JUND_CHICK	P27921 gallus gall
17	60	30.9	348	1 SXL_CERCA	O61374 ceratitis c
18	60	30.9	440	1 DCO_DROME	O76324 drosophila
19	60	30.9	497	1 FXD2_HUMAN	O60548 homo sapien
20	60	30.9	569	1 K1CJ_MOUSE	P02335 mus musculu
21	60	30.9	888	1 KLFK_MOUSE	P08923 mus musculu
22	60	30.9	1322	1 SUS_DROME	P22293 drosophila
23	59.5	30.7	391	1 SOX1_MOUSE	P53783 mus musculu
24	59	30.4	367	1 BET3_MESAU	O09029 mesocricetu
25	59	30.4	401	1 HB9_HUMAN	P50219 homo sapien
26	59	30.4	485	1 ONC2_HUMAN	O95948 homo sapien
27	59	30.4	753	1 ZIN_HUMAN	O99113 homo sapien
28	59	30.4	757	1 ECR_LUCCU	O18531 lucilia cup
29	59	30.4	4499	1 DYHA_CHLRE	O39610 chlamydomon
30	58.5	30.2	342	1 HXD9_HUMAN	P28356 homo sapien
31	58	29.9	339	1 HXD9_MOUSE	P28357 mus musculu
32	58	29.9	445	1 H3R_HUMAN	O995n1 homo sapien
33	58	29.9	476	1 EVX2_HUMAN	Q03628 homo sapien

34	58	29.9	495	1 BRN1_MOUSE	P31361 mus musculu
35	58	29.9	497	1 BRN1_RAT	Q63262 rattus norv
36	58	29.9	500	1 BRN1_HUMAN	P20264 homo sapien
37	58	29.9	517	1 Y967_TREPA	O83933 treponema p
38	58	29.9	688	1 BOMD_MOUSE	O54839 mus musculu
39	58	29.9	796	1 KF3C_RAT	O55165 rattus norv
40	58	29.9	1171	1 PHYB_ORISA	P25764 oryza sativ
41	57.5	29.6	105	1 INS_BOVIN	P01317 bos taurus
42	57.5	29.6	105	1 INS_SHEEP	P01318 ovis aries
43	57	29.4	112	1 TTFL_CAVPO	P97273 cavia porce
44	57	29.4	266	1 CANS_PIG	P04574 sus scrofa
45	57	29.4	268	1 CANS_HUMAN	P04632 homo sapien

ALIGNMENTS

RESULT 1
LAC1_NEUCR STANDARD; PRT; 619 AA.
AC P06811;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laccase precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
DE (Urishiol oxidase) (Laccase allele OR).
GN LACC.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88087214; PubMed=2961749;
RA Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
RT "Characterization of two allelic forms of Neurospora crassa laccase.
RT Amino- and carboxyl-terminal processing of a precursor.";
RL J. Biol. Chem. 263:885-896(1988).
RN [2]
RP SEQUENCE OF 379-619 FROM N.A.
RX MEDLINE=87067412; PubMed=2947240;
RA Germann U.A., Lerch K.;
RT "Isolation and partial nucleotide sequence of the laccase gene from
RT Neurospora crassa: amino acid sequence homology of the protein to
RT human ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).
CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC H(2)O.
CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14554; AAA33590.1; -;
CC EMBL; M18333; AAA33591.1; -;
CC PIR; A28523; KSNCLQ.
CC PIR; A29762; A29762.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR002355; MultiCu_oxidase2.
CC Pfam; PF00394; Cu-oxidase; 3.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
CC

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

```
CC PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
CC Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
CC KW Glycoprotein; Repeat.
CC FT SIGNAL 1 21
CC FT PROPEP 22 49
CC FT CHAIN 50 606
CC FT PROPEP 607 619
CC FT DOMAIN 84 207
CC FT DOMAIN 216 373
CC FT DOMAIN 431 566
CC FT METAL 144 144
CC FT METAL 146 146
CC FT METAL 189 189
CC FT METAL 191 191
CC FT METAL 477 477
CC FT METAL 480 480
CC FT METAL 482 482
CC FT METAL 548 548
CC FT METAL 549 549
CC FT METAL 550 550
CC FT METAL 554 554
CC FT METAL 559 559
CC FT METAL 139 139
CC FT CARBOHYD 282 282
CC FT CARBOHYD 295 295
CC FT CARBOHYD 340 340
CC FT CARBOHYD 422 422
CC FT CARBOHYD 444 444
CC SQ SEQUENCE 619 AA; 68198 MW; FDESD6D78B65048E3 CRC64;

Query Match 32.7%; Score 63.5; DB 1; Length 619;
Best Local Similarity 57.7%; Pred. No. 9.9;
Matches 15; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 7 RQWLAARAGGGGGGEGPTLRQ-W 31
   || | ||||| || || |
Db 39 RQDSQAERYGGGGGCGNSPTNRQCV 64

RESULT 2
LAC2_NEUCR STANDARD; PRT; 619 AA.
ID LAC2_NEUCR
AC P10574;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laccase precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
DE (Urishiol oxidase) (Laccase allele TS).
GN LACC.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RX MEDLINE=88087214; PubMed=2961749;
RA Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
RT "Characterization of two allelic forms of Neurospora crassa laccase.
RT Amino- and carboxyl-terminal processing of a precursor.";
RL J. Biol. Chem. 263:885-896(1988).
RC -/- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE).
CC -/- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC H(2)O.
CC -/- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -/- SUBCELLULAR LOCATION: Secreted (potential).
CC -/- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -/- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; M18334; AAA33592.1; -.
CC PIR; B28523; KSNCLT.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR002355; MultiCu_oxidase2.
CC Pfam; PF00394; Cu-oxidase; 3.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
CC PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
CC Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
CC KW Glycoprotein; Repeat.
CC FT SIGNAL 1 21
CC FT PROPEP 22 49
CC FT CHAIN 50 606
CC FT PROPEP 607 619
CC FT DOMAIN 84 207
CC FT DOMAIN 216 373
CC FT DOMAIN 431 566
CC FT METAL 144 144
CC FT METAL 146 146
CC FT METAL 189 189
CC FT METAL 191 191
CC FT METAL 477 477
CC FT METAL 480 480
CC FT METAL 482 482
CC FT METAL 548 548
CC FT METAL 549 549
CC FT METAL 550 550
CC FT METAL 554 554
CC FT METAL 559 559
CC FT METAL 139 139
CC FT CARBOHYD 282 282
CC FT CARBOHYD 295 295
CC FT CARBOHYD 340 340
CC FT CARBOHYD 422 422
CC FT CARBOHYD 444 444
CC SQ SEQUENCE 619 AA; 68120 MW; 0BB6CCDE18841145 CRC64;

Query Match 32.7%; Score 63.5; DB 1; Length 619;
Best Local Similarity 57.7%; Pred. No. 9.9;
Matches 15; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 7 RQWLAARAGGGGGGEGPTLRQ-W 31
   || | ||||| || || |
Db 39 RQDSQAERYGGGGGCGNSPTNRQCV 64

RESULT 3
YR21_TRSVR STANDARD; PRT; 201 AA.
ID YR21_TRSVR
AC P25245;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 20.2 kDa protein in RNA2.
OS Tomato ringspot virus (isolate raspberry) (Tomrsv).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
OX NCBI_TaxID=12281;
RN [1]
RX MEDLINE=91311402; PubMed=1856689;
RA Rott M.E., Tremaine J.H., Rochon D.M.;
RT "Nucleotide sequence of tomato ringspot virus RNA-2.";
RL J. Gen. Virol. 72:1505-1514(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```


CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: D12477; BAA02044.1; -
 DR PIR: JQ1094; JQ1094.
 DR HSSP: P04002; IWFA.
 KW Hypothetical protein.
 FT DOMAIN 15 22 POLY-GLY.
 FT DOMAIN 61 66 POLY-GLY.
 FT DOMAIN 144 148 POLY-GLY.
 SQ SEQUENCE 201 AA; 20194 MW; 9038506E18D7B450 CRC64;

Query Match 32.5%; Score 63; DB 1; Length 201;
 Best Local Similarity 61.5%; Pred. No. 4.1;
 Matches 16; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 13 RAGGGGGGGGIE---GPTLRQLWLA 34
 Db 13 RAGGGGGGGGKEVFKAGRTLLKVLKA 38

RESULT 4

ID NT5_RAT STANDARD; PRT; 209 AA.
 AC P34131;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurotrophin-5 precursor (NT-5) (Neurotrophic factor 5) (Neurotrophin-4)
 DE (NT-4) (Neurotrophic factor 4).
 GN NT5 OR NT4 OR NT4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92212967; PubMed=1313578;
 RA Ip N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
 RA Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
 RA Yancopoulos G.D.;
 RT "Mammalian neurotrophin-4: structure, chromosomal localization,
 RT tissue distribution, and receptor specificity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
 RN [2]

SEQUENCE FROM N.A.
 RP MEDLINE=92075279; PubMed=1742028;
 RA Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
 RA Rosenthal A.;
 RT "Neurotrophin-5: a novel neurotrophic factor that activates trk and
 RT trkB.";
 RL Neuron 7:857-866(1991).

CC -1- FUNCTION: COULD SERVE AS A TARGET-DERIVED TROPHIC FACTOR FOR
 CC SENSORY AND SYMPATHETIC NEURONS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, MUSCLE, OVARY, BRAIN,
 CC HEART, STOMACH AND KIDNEY. EXPRESSED IN BOTH EMBRYO AND ADULT
 CC TISSUES.

CC -1- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: M86742; AAA41728.1; -
 DR EMBL: S69323; AAB20548.1; -
 DR PIR: JH0504; JH0504.
 DR PIR: B42687; B42687.

DR HSSP: P34130; 1B8M.
 DR InterPro: IPR002072; NGF.
 DR Pfam: PF00243; NGF. 1.
 DR PRINTS: PR00268; NGF.
 DR PRODOM: PD002052; NGF. 1.
 DR SMART: SM00140; NGF. 1.
 DR PROSITE: PS00248; NGF. 1; 1.
 DR PROSITE: PS0270; NGF. 2; 1.
 KW Growth factor; Signal;
 FT SIGNAL 1 21
 FT PROPEP 22 79
 FT CHAIN 80 209
 FT DISULFID 96 169
 FT DISULFID 140 198
 FT DISULFID 157 200
 FT CARBOHYD 75 75
 FT CONFLICT 177 177
 SQ SEQUENCE 209 AA; 22332 MW; DF5112C05C5DB85 CRC64;

Query Match 32.2%; Score 62.5; DB 1; Length 209;
 Best Local Similarity 42.5%; Pred. No. 4.8;
 Matches 17; Conservative 2; Mismatches 12; Indels 9; Gaps 2;

QY 3 GPTLRQWL-----AARAGGG---GGGGIEGPTLRQLWLA 33
 Db 128 GSPLRQVFFETRCKAESAGEGPGVGCGRGVDRRHWS 167

RESULT 5

ID SC02_HUMAN STANDARD; PRT; 266 AA.
 AC O43819; Q9UK87;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE SC02 protein homolog, mitochondrial precursor.
 GN SC02.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Monocytes;
 RA Smink L.J., Burton J.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

[2]
 RP SEQUENCE FROM N.A., AND VARIANTS FIC LYS-140 AND PHE-225.
 RX MEDLINE=20014747; PubMed=10545952;
 RA Papadopolou L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,
 RA Sadlock J.E., Krishna S., Walker W., Selby J., Glorum D.M.,
 RA Van Coster R., Lyon G., Scalais E., Lebel R., Kaplan P., Shanske S.,
 RA De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.;
 RT "Fatal infantile cardioencephalomyopathy with COX deficiency and
 RT mutations in SC02, a COX assembly gene.";
 RL Nat. Genet. 23:333-337(1999).

CC -1- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER
 CC TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
 CC -1- TISSUE SPECIFICITY: UBUIQUITOUS.
 CC -1- DISEASE: DEFECTS IN SC02 ARE THE CAUSE OF FATAL INFANTILE
 CC CARDIOENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS
 CC CHARACTERIZED BY HYPERTROPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND
 CC GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX
 CC ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX
 CC DEFICIENCIES.

CC -1- SIMILARITY: BELONGS TO THE SC01/2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

Query Match 31.4%; Score 61; DB 1; Length 394;
 Best Local Similarity 84.6%; Pred. No. 12;
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 RAGGGGGGGGIEG 25
 | | | | | | | | | |
 DB 82 RGGGGGGGGGEEG 94

RESULT 8
 OC3N_HUMAN STANDARD; PRT; 443 AA.
 AC P20265; Q14960;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nervous-system specific octamer-binding transcription factor N-OCT 3
 DE (Brain-specific homeobox/POU domain protein 2) (BRN-2 protein)
 DE [Contains: N-OCT 5A; N-OCT 5B].
 GN POU3F2 OR BRN2 OR OTF7 OR OCT7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93181199; PubMed=8441633;
 RA Schreiber E., Tobler A., Malipiero U., Schaffner W., Fontana A.;
 RT "cDNA cloning of human N-Oct3, a nervous-system specific POU domain
 RT transcription factor binding to the octamer DNA motif.";
 RL Nucleic Acids Res. 21:253-258(1993).
 [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95380176; PubMed=7651733;
 RA Angus J., Thomson F., Murphy K., Baker E., Sutherland G.R.,
 RA Parsons P.G., Sturm R.A.;
 RT "The brn-2 gene regulates the melanocytic phenotype and tumorigenic
 RT potential of human melanoma cells.";
 RL Oncogene 11:691-700(1995).
 [3]
 RN SEQUENCE OF 280-404 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=89295573; PubMed=2739723;
 RA He X., Treacy M.N., Simmons D.M., Ingraham H.A., Swanson L.W.,
 RA Rosenfeld M.G.;
 RT "Expression of a large family of POU-domain regulatory genes in
 RT mammalian brain development.";
 RL Nature 340:35-42(1989).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
 CC RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,
 CC ('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION
 CC OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER
 CC THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II
 CC PROMOTERS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: N-OCT 3 (SHOWN HERE), N-OCT 5A
 CC AND N-OCT 5B; ARE PRODUCED BY ALTERNATIVE INITIATION.
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL
 CC CELL LINEAGE.
 CC -1- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
 CC TO CLASS-3 POU.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z11933; CAA77990.1; -;
 DR EMBL; L37868; AAB59611.1; -;
 DR PIR; S05043; S05043.
 DR PIR; S29334; S29334.
 DR HSP; P14859; I0CT.
 DR TRANSFAC; T00630; -;
 DR MIM; 600494; -;
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000327; POU.
 DR Pfam; PF000046; homeobox; 1.
 DR Pfam; PF00157; pou; 1.
 DR PRINTS; PR00028; POU DOMAIN.
 DR PRODOM; PD000583; POU; 1.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00352; POU; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00035; POU_1; 1.
 DR PROSITE; PS00465; POU_2; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 KW DNA-binding; Nuclear protein; Homeobox; Transcription regulation;
 KW Activator; Alternative initiation.
 FT CHAIN 1 443 N-OCT 3.
 FT CHAIN 181 443 N-OCT 5A.
 FT CHAIN 200 443 N-OCT 5B.
 FT INIT_MET 181 FOR N-OCT 5A.
 FT INIT_MET 200 FOR N-OCT 5B.
 FT DOMAIN 68 90 POLY-GLY.
 FT DOMAIN 125 149 POLY-GLN.
 FT DOMAIN 266 336 POU.
 FT DNA_BIND 354 413 HOMEBOX.
 FT CONFLICT 26 26 A -> G (IN REF. 2).
 SQ SEQUENCE 443 AA; 46921 MW; 2CAC852328334A66 CRC64;

Query Match 31.4%; Score 61; DB 1; Length 443;
 Best Local Similarity 66.7%; Pred. No. 13;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 QWLARAGGGGGGG 22
 | | | | | | | | | |
 DB 60 QWITALSHGGGGGG 74

RESULT 9
 OC3N_MOUSE STANDARD; PRT; 445 AA.
 AC P31360;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Nervous-system specific octamer-binding transcription factor N-OCT 3
 DE (Brain-specific homeobox/POU domain protein 2) (BRN-2 protein).
 GN POU3F2 OR OTF7 OR BRN2 OR BRN-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9228768; PubMed=1565620;
 RA Hara Y., Rovescalli C., Kim Y., Nirenberg M.;
 RT "Structure and evolution of four POU domain genes expressed in mouse
 RT brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3280-3284(1992).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
 CC RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,
 CC ('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION
 CC OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER
 CC THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II
 CC PROMOTERS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL
 CC CELL LINEAGE.

```

CC CC -!- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
CC CC TO CLASS-3 POU.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88300; AAA39961.1;
CC FIRM; S31224; S31224.
CC HSP; P14859; LOC.
CC MGD; MGI:101895; Pou3f2.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC Pfam; PF00157; pou; 1.
CC PRINTS; PR00028; POUDOMAIN.
CC ProDom; PD000583; POU; 1.
CC SMART; SM00389; HOX; 1.
CC SMART; SM00352; POU; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC PROSITE; PS00035; POU_1; 1.
CC PROSITE; PS00465; POU_2; 1.
CC DNA-binding; Nuclear protein; Homeobox; Transcription regulation;
KW Activator.
FT DOMAIN 68 90 POLY-GLY.
FT DOMAIN 125 149 POLY-GLN.
FT DOMAIN 268 338 POU.
FT DNA_BIND 356 415 HOMEBOX.
SQ SEQUENCE 445 AA; 67149 MW; 1A47F10950EECE8A CRC64;

Query Match 31.4%; Score 61; DB 1; Length 445;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 OWLAARAGGGGGGG 22
DB 60 QWITALSHGGGGGG 74
||:|:|||||||

RESULT 10
ID CNAL_DROME STANDARD; PRT; 584 AA.
AC P12252;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE CAMP-dependent 3',5'-cyclic phosphodiesterase (EC 3.1.4.17) (Learning/
DE memory process protein).
DE DUNCE OR DNC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS.
RC STRAIN=CANTON-S;
RX MEDLINE=92085274; PubMed=1660926;
RA Qiu Y.H., Chen C.-N., Malone T., Richter L., Beckendorf S.K.,
RA Davis R.L.;
RT "Characterization of the memory gene dunce of Drosophila
RT melanogaster.";
RL J. Mol. Biol. 222:553-565(1991).
RN [2]
RP SEQUENCE OF 223-584 FROM N.A.
RX MEDLINE=87092243; PubMed=3025834;
RA Chen C.-N., Denome S., Davis R.L.;
RA "Molecular analysis of cDNA clones and the corresponding genomic
RT

```

```

RT RT coding sequences of the Drosophila dunce+ gene, the structural gene
RT for cAMP phosphodiesterase.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9313-9317(1986).
CC -!- CATALYTIC ACTIVITY: ADENOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
CC ADENOSINE 5'-PHOSPHATE.
CC -!- PATHWAY: CYCLIC NUCLEOTIDE METABOLISM.
CC -!- SUBUNIT: MONOMER.
CC -!- ALTERNATIVE PRODUCTS: DIFFERENT FORMS ARE GENERATED BY THE USE OF
CC DIFFERENT TRANSCRIPTION START SITES AND SPLICING PATTERNS.
CC -!- DISEASE: MUTATION OF DUNCE PRODUCES FEMALE FLIES THAT ARE STERILE.
CC -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X55167; CAA38960.1;
CC EMBL; X55168; CAA38960.1; JOINED.
CC EMBL; X55169; CAA38960.1; JOINED.
CC EMBL; X55170; CAA38960.1; JOINED.
CC EMBL; X55171; CAA38960.1; JOINED.
CC EMBL; X55172; CAA38960.1; JOINED.
CC EMBL; X55173; CAA38960.1; JOINED.
CC EMBL; X55174; CAA38960.1; JOINED.
CC EMBL; X55175; CAA38960.1; JOINED.
CC EMBL; M14982; AAC34201.1; JOINED.
CC EMBL; M14978; AAC34201.1; JOINED.
CC EMBL; M14979; AAC34201.1; JOINED.
CC EMBL; M14980; AAC34201.1; JOINED.
CC EMBL; M14981; AAC34201.1; JOINED.
CC PIR; A26651; A26651.
CC FlyBase; FBgn0000479; dnc.
CC InterPro; IPR003607; HDC.
CC InterPro; IPR002073; PDEase.
CC Pfam; PF00233; PDEase; 1.
CC PRINTS; PR00387; PDIESTERASE1.
CC SMART; SM00471; HDC; 1.
CC PROSITE; PS00126; PDEASE_I; 1.
CC Hydrolase; CAMP; Alternative splicing.
FT DOMAIN 305 310 PART OF CAMP BINDING SITE (BY SIMILARITY
FT PIR; A26651; A26651. TO MAMMALIAN REGULATORY SUBUNIT OF TYPE 2
FT InterPro; IPR003607; HDC. CAMP DEPENDENT PROTEIN KINASE).
FT PIR; A26651; A26651. THR-RICH.
FT DOMAIN 542 551 GLY-RICH.
FT DOMAIN 559 567
SQ SEQUENCE 584 AA; 64875 MW; 99239BE33C620501 CRC64;

Query Match 31.4%; Score 61; DB 1; Length 584;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGIEGP 26
DB 555 ALRAGGGGGGGGMAP 570
||:|||||||

RESULT 11
ID KICJ_HUMAN STANDARD; PRT; 593 AA.
AC P13645;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type I cytoskeletal 10 (cytokeratin 10) (CK 10).
GN KRT10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

RN [1] SEQUENCE FROM N.A.
RP MEDLINE-89125611; PubMed-2464696;
RX Rieger M., Franke W.W.;
RA "Identification of an orthologous mammalian cytokeratin gene. High
RT degree of intron sequence conservation during evolution of human
RT cytokeratin 10.";
RL J. Mol. Biol. 204:841-856(1988).
RN [2]
RP SEQUENCE OF 130-593 FROM N.A.
RX MEDLINE-88122104; PubMed-2448602;
RA Darmon M.Y., Semat A., Darmon M.C., Vasseur M.;
RT "Sequence of a cDNA encoding human keratin No 10 selected according
RT to structural homologies of keratins and their tissue-specific
RT expression.";
RL Mol. Biol. Rep. 12:277-283(1987).
RN [3]
RP SEQUENCE OF 197-593 FROM N.A.
RX MEDLINE-92339897; PubMed-1378806;
RA Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P.,
RA Kisselev L.L.;
RT "Exons I and VII of the gene (Ker10) encoding human keratin 10
RT undergo structural rearrangements within repeats.";
RL Gene 116:245-251(1992).
RN [4]
RP SEQUENCE OF 180-184 AND 577-589.
RC TISSUE-Keratinocytes;
RX MEDLINE-93162043; PubMed-1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
RN [5]
RP VARIANT EHK HIS-156.
RX MEDLINE-92386600; PubMed-1381287;
RA Cheng J., Syder A.J., Yu Q.-C., Letal A., Paller A.S., Fuchs E.;
RT "The genetic basis of epidermolytic hyperkeratosis: a disorder of
RT differentiation-specific epidermal keratin genes.";
RL Cell 70:811-819(1992).
RN [6]
RP VARIANTS.
RX MEDLINE-92141228; PubMed-1371013;
RA Korge B.P., Gan S.-Q., McBride O.W., Mischke D., Steinert P.M.;
RT "Extensive size polymorphism of the human keratin 10 chain resides in
RT the C-terminal V2 subdomain due to variable numbers and sizes of
RT glycine loops.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).
RN [7]
RP VARIANTS EHK HIS-156 AND SER-161.
RX MEDLINE-92376531; PubMed-1380725;
RA Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT hyperkeratosis.";
RL Science 257:1128-1130(1992).
RN [8]
RP VARIANTS EHK HIS-154; CYS-156; HIS-156; ASP-160 AND GLN-442.
RX MEDLINE-94136477; PubMed-7508181;
RA Chipev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L.,
RA Compton J.G., Bale S.J.;
RT "Preferential sites in keratin 10 that are mutated in epidermolytic
RT hyperkeratosis.";
RL Am. J. Hum. Genet. 54:179-190(1994).
RN [9]
RP VARIANTS EHK ARG-150; CYS-156 AND GLU-439, AND VARIANT SER-126.
RX MEDLINE-94216497; PubMed-7512983;
RA Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;
RT "Genetic mutations in the K1 and K10 genes of patients with
RT epidermolytic hyperkeratosis. Correlation between location and
RT disease severity.";
RL J. Clin. Invest. 93:1533-1542(1994).
RN [10]
RP VARIANTS EHK ASN-160.
RX MEDLINE-94117868; PubMed-7507150;
RA Rothnagel J.A., Longley M.A., Holder R.A., Kuster W., Roop D.R.;
RT "Prenatal diagnosis of epidermolytic hyperkeratosis by direct gene
RT sequencing.";
RL J. Invest. Dermatol. 102:13-16(1994).
RN [11]
RP VARIANTS EHK PRO-156 AND SER-156.
RX MEDLINE-94117870; PubMed-7507152;
RA McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,
RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA Morley S.M.;
RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT congenital ichthyosiform erythroderma (BCIE).";
RL J. Invest. Dermatol. 102:24-30(1994).
RN [12]
RP VARIANT EHK THR-150.
RX MEDLINE-95059228; PubMed-7526210;
RA Paller A.S., Syder A.J., Chan Y.-M., Yu Q.-C., Hutton M.E., Tadini G.,
RA Fuchs E.;
RT "Genetic and clinical mosaicism in a type of epidermal nevus.";
RL New Engl. J. Med. 331:1408-1415(1994).
RN [13]
RP VARIANT AEI THR-446.
RX MEDLINE-99072665; PubMed-9856845;
RA Suga Y., Duncan K.O., Heald P.W., Roop D.R.;
RT "A novel helix termination mutation in keratin 10 in annular
RT epidermolytic ichthyosis, a variant of bullous congenital
RT ichthyosiform erythroderma.";
RL J. Invest. Dermatol. 111:1220-1223(1998).
RN [14]
RP VARIANT EHK SER-160.
RX MEDLINE-99215719; PubMed-10201536;
RA Arin M.J., Longley M.A., Anton-Lamprecht I., Kurze G., Huber M.,
RA Hohl D., Rothnagel J.A., Roop D.R.;
RT "A novel substitution in keratin 10 in epidermolytic hyperkeratosis.";
RL J. Invest. Dermatol. 112:506-508(1999).
RN [15]
RP SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.
CC [1- TISSUE SPECIFICITY: SEEN IN ALL SUPRABASAL CELL LAYERS INCLUDING
CC STRATUM CORNEUM.
CC [1- POLYMORPHISM: A NUMBER OF ALLELES ARE KNOWN THAT MAINLY DIFFER IN
CC THE GLY-RICH REGION (POSITIONS 490-560).
CC [1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF EPIDERMOLYTIC
CC HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC ICHTHYOSIFORM ERYTHRODERMA (BCIE)); A HEREDITARY SKIN DISORDER
CC CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC CORNEUM. AT BIRTH, AFFECTED INDIVIDUALS USUALLY PRESENT WITH
CC REDNESS, BLISTERS AND SUPERFICIAL EROSIONS DUE TO CYTOLYSIS.
CC WITHIN A FEW WEEKS, THE ERYTHRODERMA AND BLISTER FORMATION
CC DIMINISH AND HYPERKERATOSIS DEVELOP. TRANSMISSION IS AUTOSOMAL
CC DOMINANT, BUT MOST CASES ARE SPORADIC.
CC [1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF ANNULAR EPIDERMOLYTIC
CC ICHTHYOSIS (AEI), A DISTINCT PHENOTYPIC VARIANT OF EPIDERMOLYTIC
CC HYPERKERATOSIS. IT RESEMBLES CLINICAL AND HISTOLOGIC FEATURES OF
CC BOTH EPIDERMOLYTIC HYPERKERATOSIS AND ICHTHYOSIS BULLOSA OF
CC SIEMENS.
CC [1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II
CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].
CC [1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC [1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY IN
CC POSITIONS 513 TO 555.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14487; CAA32649.1; -

DR EMBL; M19156; AAA59468.1; -;
 DR EMBL; M77663; AAA59199.1; -;
 DR EMBL; L20218; AAB59438.1; -;
 DR EMBL; L20219; AAB59439.1; -;
 DR PIR; S02158; KRHU0.
 DR Aarhus/Chent-2DPAGE; 7405; IEF.
 DR MIM; 148080; -;
 DR MIM; 113800; -;
 DR InterPro; IPR001664; IF.
 DR InterPro; IPR002957; Keratin_1.
 DR Pfam; PF00038; filament; 1.
 DR PRINTS; PR01248; TYPE1KERATIN.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Keratin; Disease mutation;
 KW Polymorphism.
 FT DOMAIN 1 145 HEAD.
 FT DOMAIN 146 456 ROD.
 FT DOMAIN 457 593 TAIL.
 FT DOMAIN 146 181 COIL 1A.
 FT DOMAIN 182 202 LINKER 1.
 FT DOMAIN 203 294 COIL 1B.
 FT DOMAIN 295 317 LINKER 12.
 FT DOMAIN 318 456 COIL 2.
 FT DOMAIN 6 144 GLY/PHE/SER-RICH.
 FT DOMAIN 451 590 GLY/SER-RICH.
 FT VARIANT 126 126 G -> S.
 FT VARIANT 150 150 /FTid=VAR_010505.
 FT VARIANT 150 150 M -> R (IN EHK).
 FT VARIANT 150 150 /FTid=VAR_010506.
 FT VARIANT 154 154 M -> T (IN EHK).
 FT VARIANT 156 156 N -> H (IN EHK).
 FT VARIANT 156 156 R -> H (IN EHK).
 FT VARIANT 156 156 /FTid=VAR_003826.
 FT VARIANT 156 156 /FTid=VAR_003827.
 FT VARIANT 156 156 R -> C (IN EHK).
 FT VARIANT 156 156 /FTid=VAR_003828.
 FT VARIANT 156 156 R -> P (IN EHK).
 FT VARIANT 156 156 /FTid=VAR_003829.
 FT VARIANT 156 156 R -> S (IN EHK).
 FT VARIANT 160 160 /FTid=VAR_003830.
 FT Y -> D (IN EHK; SEVERE PHENOTYPE).
 FT /FTid=VAR_003831.
 Query Match 31.4%; Score 61; DB 1; Length 593;
 Best Local Similarity 52.6%; Pred. No. 17;
 Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 OY 7 ROWLAARAGGGGGGGIEG 25
 Db 9 KHYSSRSRGGGGGGCGG 27
 RESULT 12
 ID MYSC_ACACA STANDARD; PRT; 1168 AA.
 AC P10569;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin IC heavy chain.
 GN MIC.
 OS Acanthamoeba castellanii (Amoeba).
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.
 OX NCBI_TaxID=5755;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=68016163; PubMed=3477803;
 RA Jung G., Korn E.D., Hammer J.A. III;
 RA "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
 RT and non-myosin-like sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).
 EN [2]

RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=86259656; PubMed=3014500;
 RA Hammer J.A. III, Jung G., Korn E.D.;
 RT "Genetic evidence that Acanthamoeba myosin I is a true myosin."
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659(1986).
 RN [3]
 RP PHOSPHORYLATION SITE.
 RX MEDLINE=90037074; PubMed=2530230;
 RA Brzeska H., Lynch T.J., Martin B., Korn E.D.;
 RT "The localization and sequence of the phosphorylation sites of
 Acanthamoeba myosins I. An improved method for locating the
 RT phosphorylated amino acid."
 RL J. Biol. Chem. 264:19340-19348(1989).
 CC -|- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
 CC -|- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
 CC CHAIN AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
 CC -|- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
 CC MYOSIN I CAN THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND VICE
 CC VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
 CC THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
 CC OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
 CC -|- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
 CC MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
 CC -|- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -|- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; J02974; AAA27707.1; -;
 DR PIR; A33891; MWXIC.
 DR HSP; P08799; LLVK.
 DR InterPro; IPR001452; SH3.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00242; MYSC; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.
 KW DOMAIN 1 670 MYOSIN HEAD-LIKE.
 FT DOMAIN 671 922 TAIL HOMOLGY REGION 1 (TH.1).
 FT DOMAIN 923 975 GLY/PRO/ALA-RICH (TH.2).
 FT DOMAIN 976 1035 SH3.
 FT DOMAIN 1036 1168 GLY/PRO/ALA-RICH (TH.2).
 FT NP_BIND 101 108 ATP (POTENTIAL).
 FT MOD_RES 311 311 PHOSPHORYLATION.
 SQ SEQUENCE 1168 AA; 127309 MW; D07084B373A37A32 CRC64;
 Query Match 31.4%; Score 61; DB 1; Length 1168;
 Best Local Similarity 60.0%; Pred. No. 31;
 Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 OY 8 OWLAARAGGGGGGGIEGPT 27
 Db 920 QILGAKGGGGGGGRGGRGPS 939
 RESULT 13
 ID PHVB_SORBI STANDARD; PRT; 1178 AA.
 AC P93527;
 DT 16-OCT-2001 (Rel. 40, Created)


```

or send an email to license@isb-sib.ch).
-----
CC EMBL; M86528; AAA60154.1; -.
CC DR PIR; JH0503; JH0503.
CC DR PIR; A42687; A42687.
CC DR PDB; 1B8M; 09-FEB-99.
CC DR PDB; 1B98; 26-FEB-99.
CC DR MIM; 162662; -.
CC DR InterPro; IPR002072; NGF.
CC DR Pfam; PF00243; NGF; 1.
CC DR PRINTS; PR00268; NGF.
CC DR ProDom; PD002052; NGF; 1.
CC DR SMART; SM00140; NGF; 1.
CC DR PROSITE; PS00248; NGF_1; 1.
CC DR PROSITE; PS02270; NGF_2; 1.
CC KW Growth factor; Signal; 3D-structure.
FT SIGNAL 1 24
FT PROPEP 25 80
FT CHAIN 81 210
FT DISULFID 97 170
FT DISULFID 141 199
FT DISULFID 158 201
FT CARBOHYD 76 76
SQ SEQUENCE 210 AA; 22426 MW; DBC6A30195E139AD CRC64;
Query Match 31.2%; Score 60.5; DB 1; Length 210;
Best Local Similarity 35.0%; Pred. No. 7.6;
Matches 14; Conservative 3; Mismatches 14; Indels 9; Gaps 1;
QY 3 GPTLRQWL-----ARAGGGGGGGGIEGPTLRQWLA 33
DB 129 GSPLOYFFETRCKADNAEEGGPGAGGGCGVDRRHWS 168
RESULT 15
KLTK_HUMAN
ID KLTK_HUMAN STANDARD; PRT; 864 AA.
AC P29376;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leukocyte tyrosine kinase receptor precursor (EC 2.7.1.112) (Protein
  tyrosine kinase-1).
GN LTK OR TYKL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93296146; PubMed=7685902;
RA Toyoshima H., Kozutsumi H., Maru Y., Hagiwara K., Furuya A.,
RA Miho H., Hanai N., Takaku F., Yazaki Y., Hirai H.;
RT "Differentially spliced cDNAs of human leukocyte tyrosine kinase
  receptor tyrosine kinase predict receptor proteins with and without a
  tyrosine kinase domain and a soluble receptor protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5404-5408(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92007735; PubMed=1655406;
RA Krolewski J.J., Dalla-Favera R.;
RT "The ltk gene encodes a novel receptor-type protein tyrosine kinase.";
RL EMBO J. 10:2911-2919(1991).
RN [3]
RP SEQUENCE OF 416-864 FROM N.A.
RX MEDLINE=90206632; PubMed=2320375;
RA Maru Y., Hirai H., Takaku F.;
RT "Human ltk: gene structure and preferential expression in human
  leukemic cells.";
RL Oncogene Res. 5:199-204(1990).
CC -!- FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS NOT KNOWN. IT IS
  PROBABLY A RECEPTOR WITH A TYROSINE-PROTEIN KINASE ACTIVITY.
-----
or send an email to license@isb-sib.ch).
-----
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
  tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: LAMBDA P1, LAMBDA P2
  (SHOWN HERE) AND LAMBDA P3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
  PROTEIN KINASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
CC EMBL; D16105; BAA03679.1; -.
CC EMBL; X60702; CAA43113.1; -.
CC EMBL; X52213; CAA36460.1; -.
CC PIR; S17452; S17452.
CC HSSP; P00523; 2PTK.
CC MIM; 151520; -.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002011; Receptor_tyr_kin_II.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; Transmembrane; ATP-binding;
  Phosphorylation; Receptor; Glycoprotein; Alternative splicing;
  Signal.
KW SIGNAL
FT CHAIN 1 16
FT DOMAIN 17 864
FT TRANSMEM 425 449
FT DOMAIN 450 864
FT DOMAIN 510 786
FT NP_BIND 516 524
FT BINDING 544 544
FT ACT_SITE 643 643
FT MOD_RES 676 676
FT CARBOHYD 257 257
FT CARBOHYD 380 380
FT CARBOHYD 412 412
FT VARSPLIC 170 170
FT VARSPLIC 171 864
FT VARSPLIC 448 448
FT VARSPLIC 449 864
FT CONFLICT 42 42
FT CONFLICT 220 220
FT CONFLICT 274 334
FT CONFLICT 449 449
FT CONFLICT 652 654
SQ SEQUENCE 864 AA; 91653 MW; 97143DD57684A657 CRC64;
Query Match 31.2%; Score 60.5; DB 1; Length 864;
Best Local Similarity 63.6%; Pred. No. 26;
Matches 14; Conservative 1; Mismatches 2; Indels 5; Gaps 2;
QY 2 EG-PTLRQWLAARAGGGGGGGG 22
DB 196 EGVPGRRW-----AGGGGGGGG 213
RESULT 16
JUND_CHTCK
ID JUND_CHTCK STANDARD; PRT; 323 AA.

```

AC p27921;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 GN Transcription factor Jun-D.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92019832; PubMed=1923529;
 RA Hartl M., Hutchins J.T., Vogt P.K.;
 RA "The chicken JunD gene and its product.";
 RL Oncogene 6:1623-1631(1991).
 CC -1- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. JUN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X60063; CAA42665.1; -
 DR PIR; S20099; S20099.
 DR HSP; P05412; IFOS.
 DR TRANSFAC; T02196; -
 DR InterPro; IPR002112; Leuzip_Jun.
 DR InterPro; IPR001871; bzip.
 DR Pfam; PF00170; bzip; 1.
 DR PRINTS; PR00043; LEUZIPPJUN.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 DR Transcription regulation; DNA-binding; Activator; Nuclear protein.
 KW DOMAIN 59 67 POLY-ALA.
 FT DOMAIN 155 166 POLY-GLY.
 FT DNA_BIND 242 266 BASIC MOTIF.
 FT DOMAIN 270 298 LEUCINE-ZIPPER.
 SQ SEQUENCE 323 AA; 33205 MW; A7F6D21A97DBB676 CRC64;
 Query Match 30.9%; Score 60; DB 1; Length 323;
 Best Local Similarity 72.2%; Pred. No. 12;
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 11 AARAGGGGGGGGIEGPTL 28
 Db 151 AAAAGGGGGGGGGGEL 168
 |||||
 RESULT 17
 ID SXL_CERCA STANDARD; PRT; 348 AA.
 AC O61374;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sex-lethal protein homolog (CCSXL).
 GN SXL.
 OS Ceratitis capitata (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Tephritidae; Tephritidae; Ceratitis.
 OX NCBI_TaxID=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-BENAKIO.
 RC MEDLINE=98171464; PubMed=9502730;
 RX

RA Saccone G., Peluso I., Artiaco D., Giordano E., Bopp D., Polito L.C.;
 RT "The Ceratitis capitata homologue of the Drosophila sex-determining
 RT gene Sex-lethal is structurally conserved, but not sex-specifically
 RT regulated.";
 RL Development 125:1495-1500(1998).
 CC -1- FUNCTION: UNKNOWN; APPARENTLY NOT INVOLVED IN SOMATIC SEX
 CC DETERMINATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS: ADULT-SPECIFIC ISOFORMS
 CC AL, A2, A3, A4, AND EMBRYO-SPECIFIC ISOFORMS E1, E2 AND E3 (SHOWN
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYOS OF BOTH SEXES. ALSO
 CC EXPRESSED IN THE PROGENITOR CELLS OF THE GERMLINE.
 CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF026145; AAC38968.1; -
 DR HSP; P19339; ISXL.
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; rrm; 2.
 DR PRINTS; PR00961; HUDSLRNA.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS01012; RRM; 2.
 DR PROSITE; PS00030; Repeat; Nuclear
 KW RNA-binding; Repeat; Nuclear
 FT DOMAIN 1 27 GLY/ASN-RICH DOMAIN.
 FT DOMAIN 110 188 RNA-BINDING (RRM) 1.
 FT DOMAIN 196 276 RNA-BINDING (RRM) 2.
 FT DOMAIN 68 75 POLY-GLY.
 FT DOMAIN 95 99 POLY-GLY.
 FT DOMAIN 293 311 POLY-GLY.
 FT DOMAIN 312 316 POLY-PRO.
 FT VARSPLIC 37 44 MISSING (IN ISOFORM A1).
 SQ SEQUENCE 348 AA; 37188 MW; CABA3DA5C2C8874A CRC64;
 Query Match 30.9%; Score 60; DB 1; Length 348;
 Best Local Similarity 83.3%; Pred. No. 13;
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 15 GGGGGGGGIEGP 26
 Db 301 GGGGGGGGGG 312
 |||||
 RESULT 18
 ID DCO_DROME STANDARD; PRT; 440 AA.
 AC O76324;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Discs overgrown protein kinase (BC 2.7.1.-) (Double-time protein).
 GN DCO OR DBT.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98337186; PubMed=9674431;
 RA Kloss B., Price J.L., Saez L., Blau J., Rothenfluh A., Wesley C.S.,
 RA Young M.W.;
 RT "The Drosophila clock gene double-time encodes a protein closely
 RT related to human casein kinase I epsilon.";
 RL Cell 94:97-107(1998).
 RL

[2]
 RN MUTAGENESIS, AND FUNCTION.
 RX MEDLINE-98337187; PubMed-9674430;
 RA Price J.L., Blau J., Rothenfluh A., Abodeely M., Kloss B., Young M.W.;
 RT "Double-time is a novel Drosophila clock gene that regulates PERIOD
 RL protein accumulation.";
 RL Cell 94:83-95(1998).
 CC -!- FUNCTION: INVOLVED IN CIRCADIAN RHYTHMS, VIABILITY AND MOLECULAR
 CC OSCILLATIONS OF THE CLOCK GENES PERIOD (PER) AND TIMELESS (TIM).
 CC DBT REDUCES THE STABILITY AND THUS THE ACCUMULATION OF MONOMERIC
 CC PER PROTEINS, PROBABLY THROUGH PHOSPHORYLATION. NO EVIDENT
 CC CIRCADIAN OSCILLATION IS DETECTED IN HEAD.
 CC -!- SUBUNIT: FORMS A COMPLEX WITH PER.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN PHOTORECEPTOR CELLS OF THE EYES
 CC AS WELL AS IN THE REGION SITUATED BETWEEN THE OPTIC LOBE AND THE
 CC CENTRAL BRAIN.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CASEIN KINASE I SUBFAMILY. COULD BE THE ORTHOLOG OF CKI-EPSILON.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AF055583; AAC39134.1; -;
 DR HSSP; P40233; 1CSN.
 DR FlyBase; FBgn002413; dco.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Biological rhythms; Transferase; Serine/threonine-protein kinase;
 KW ATP-binding.
 FT DOMAIN 9 277 PROTEIN KINASE.
 FT NP_BIND 15 28 ATP (BY SIMILARITY).
 FT BINDING 35 33 ATP (BY SIMILARITY).
 FT ACT_SITE 128 128 BY SIMILARITY.
 FT DOMAIN 319 332 POLY-ALA.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 351 POLY-GLY.
 FT DOMAIN 414 426 POLY-GLY.
 FT DOMAIN 430 437 POLY-GLY.
 FT MUTAGEN 47 47 P->S; IN DBTS; SHORTENS THE BEHAVIORAL
 FT MUTAGEN 80 80 PERIOD.
 FT MUTAGEN 80 80 M->I; IN DBTL; LENGTHENS THE BEHAVIORAL
 FT PERIOD.
 SQ SEQUENCE 440 AA; 48073 MW; B875891D5747391D CRC64;
 Query Match 30.9%; Score 60; DB 1; Length 440;
 Best Local Similarity 55.0%; Pred. No. 16;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 4 PTLRWLAARAGGGGGGGI 23
 Db 403 PERRPSIRMRQGGGGGGV 422

 RESULT 19
 FXD2_HUMAN
 ID FXD2_HUMAN STANDARD; PRT; 497 AA.
 AC O60548;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Forkhead box protein D2 (Forkhead-related protein FKHL17) (Forkhead-
 DE related transcription factor 9) (FREAC-9).
 GN FOXD2 OR FKHL17 OR FREAC9.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98066765; PubMed-9403061;
 RA Ernstsson S., Betz R., Lagercrantz S., Larsson C., Ericksson S.,
 RA Cederberg A., Carlsson P., Enerbaeck S.;
 RT "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
 RT expressed human forkhead gene that maps to chromosome 1p32-p34.";
 RL Genomics 46:78-85(1997).
 RN [2]
 RP REVISIONS.
 RA Enerbaeck S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
 CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AF042832; AAC15421.1; -;
 DR HSSP; Q63245; 2HFH.
 DR TRANSFAC; T02485; -;
 DR MIN; 602211; -;
 DR InterPro; IPR001766; Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR SMART; SMO0339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS50039; FORK_HEAD_3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation.
 FT DOMAIN 90 94 POLY-ALA.
 FT DOMAIN 101 104 POLY-ALA.
 FT DOMAIN 126 217 FORK-HEAD.
 FT DOMAIN 247 250 POLY-ALA.
 FT DOMAIN 296 306 POLY-ALA.
 FT DOMAIN 398 409 POLY-GLY.
 FT DOMAIN 421 426 POLY-GLY.
 FT DOMAIN 442 445 POLY-ALA.
 SQ SEQUENCE 497 AA; 49007 MW; EAAF498D216BE019 CRC64;
 Query Match 30.9%; Score 60; DB 1; Length 497;
 Best Local Similarity 66.7%; Pred. No. 18;
 Matches 14; Conservative 0; Mismatches 5; Indels 2; Gaps 1;
 QY 4 PT--LRQWLAARAGGGGGGG 22
 Db 385 PTLRLQGLKTDAGGGAGGG 405

 RESULT 20
 KICJ_MOUSE
 ID KICJ_MOUSE STANDARD; PRT; 569 AA.
 AC P02535; P08731;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Keratin, type I cytoskeletal 10 (Cytokeratin 10) (56 kDa cytokeratin)
 DE (Keratin, type I cytoskeletal 59 kDa).
 GN KRT10 OR KRT1-10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RN SEQUENCE FROM N.A.
RP MEDLINE=85207552; PubMed=2581944;
RX Krieg T.M., Schafer M.P., Cheng C.K., Filipula D., Flaherty P.,
RT "Organization of a type I keratin gene. Evidence for evolution of
RT intermediate filaments from a common ancestral gene.";
RL J. Biol. Chem. 260:5867-5870(1985).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=83192464; PubMed=6188955;
RX Steinert P.M., Rice R.H., Roop D.R., Trus B.L., Steven A.C.;
RT "Complete amino acid sequence of a mouse epidermal keratin subunit
RT and implications for the structure of intermediate filaments.";
RL Nature 302:794-800(1983).
CC -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L00193; AAA39391.1; .
DR EMBL; M10081; AAA39391.1; JOINED.
DR EMBL; V00830; CAA24214.1; .
DR PIR; A02940; KMSL1.
DR PIR; S07330; S07330.
DR HSP; P10968; IWGC.
DR MGD; MGI:96685; Krt1-10.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
RW Intermediate filament; Coiled coil; Keratin.
FT INIT_MET 0 0
FT DOMAIN 1 142 HEAD.
FT DOMAIN 143 453 ROD.
FT DOMAIN 454 569 TAIL.
FT DOMAIN 143 178 TAIL.
FT DOMAIN 179 199 LINKER 1A.
FT DOMAIN 200 291 COIL 1B.
FT DOMAIN 292 314 LINKER 12.
FT DOMAIN 315 453 COIL 2.
FT SITE 395 395 STUTTER.
FT DOMAIN 452 564 GLY/SER-RICH.
FT CONFLICT 5 5 S -> C (IN REF. 2).
FT CONFLICT 24 24 S -> F (IN REF. 2).
FT CONFLICT 28 28 S -> F (IN REF. 2).
FT CONFLICT 38 38 Y -> L (IN REF. 2).
FT CONFLICT 41 41 E -> G (IN REF. 2).
FT CONFLICT 104 105 AG -> GS (IN REF. 2).
FT CONFLICT 110 110 MISSING (IN REF. 2).
FT CONFLICT 121 122 SY -> GC (IN REF. 2).
FT CONFLICT 137 137 S -> G (IN REF. 2).
FT CONFLICT 148 148 Q -> R (IN REF. 2).
FT CONFLICT 178 187 WYKHHGSSQ -> VVREARQLKP (IN REF. 2).
FT CONFLICT 263 268 KSDLEM -> QSVLEL (IN REF. 2).
FT CONFLICT 284 284 H -> L (IN REF. 2).
FT CONFLICT 353 353 E -> A (IN REF. 2).
FT CONFLICT 394 399 EGRYCV -> VESLLR (IN REF. 2).
FT CONFLICT 508 514 GSHGGS -> CGGRGG (IN REF. 2).
FT CONFLICT 523 523 S -> G (IN REF. 2).
FT CONFLICT 531 531 H -> R (IN REF. 2).
FT CONFLICT 534 534 S -> G (IN REF. 2).

FT CONFLICT 543 543 S -> G (IN REF. 2).
FT CONFLICT 547 548 GQ -> RR (IN REF. 2).
FT CONFLICT 555 556 KS -> SGT (IN REF. 2).
SQ SEQUENCE 569 AA; 57711 MW; EEC59D4D8FFE484D CRC64;
Query Match 30.9%; Score 60; DB 1; Length 569;
Best Local Similarity 43.5%; Pred. No. 21;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
QY 7 ROWLAARAGGGGGGGTETPLR 29
Db 9 KQFSRRSGGGGGGVRVSSTR 31
RESULT 21
KLTK_MOUSE STANDARD; PRT; 888 AA.
ID KLTK_MOUSE
AC P08923;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Leukocyte tyrosine kinase receptor precursor (EC 2.7.1.112).
GN LTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Snijders A.J., Haase V.H., Bernards A.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 252-270 AND 332-888 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=90291994; PubMed=2357970;
RA Bernards A., de la Monte S.;
RT "The ltk receptor tyrosine kinase is expressed in pre-B lymphocytes
RT and cerebral neurons and uses a non-AUG translational initiator.";
RL EMBO J. 9:2279-2287(1990).
RN [3]
RP SEQUENCE OF 252-270 AND 332-888 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=92115335; PubMed=1662793;
RA Haase V.H., Snijders A.J., Cooke S.M., Teng M.N., Kaul D.,
RA le Beau M.M., Bruns G.A., Bernards A.;
RT "Alternatively spliced ltk mRNA in neurons predicts a receptor with a
RT larger putative extracellular domain.";
RL Oncogene 6:2319-2325(1991).
RN [4]
RP SEQUENCE OF 217-270 AND 332-888 FROM N.A.
RX MEDLINE=88232962; PubMed=2836739;
RA Ben-Neriah Y., Bauskin A.R.;
RT "Leukocytes express a novel gene encoding a putative transmembrane
RT protein-kinase devoid of an extracellular domain.";
RL Nature 333:672-676(1988).
CC -!- FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS NOT KNOWN, IT IS
CC PROBABLY A RECEPTOR WITH A TYROSINE-PROTEIN KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: TISSUE-SPECIFIC ALTERNATIVE SPLICING
CC PRODUCES VARIANTS WITH SHORTER EXTRACELLULAR DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; M90470; AAA39451.1;
DR EMBL; X52621; CAA36848.1; ALT_SEQ.
DR EMBL; X07984; CAA30793.1; ALT_INIT.
DR PIR; S00904; S00904.
DR PIR; S12792; S12792.
DR HSP; P11362; LFGR.
DR MGI; 96840; Ltk.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002011; Receptor_tyr_kin_II.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase.1
DR PRINTS; PR00109; TYRKINASE.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferrase; Tyrosine-protein kinase; Transmembrane; ATP-binding;
KW Phosphorylation; Receptor; Glycoprotein; Signal; Alternative splicing.
FT SIGNAL 1 16
FT CHAIN 17 888
FT DOMAIN 17 421
FT TRANSMEM 422 446
FT DOMAIN 447 888
FT DOMAIN 506 782
FT NP_BIND 512 520
FT BINDING 540 540
FT ACT_SITE 639 639
FT MOD_RES 672 672
FT CARBOHYD 377 377
FT CARBOHYD 409 409
FT VARSPLOC 271 331
FT CONFLICT 789 789
FT CONFLICT 875 875
FT CONFLICT 888 888
FT SEQUENCE 888 AA; 94436 MW; 3FFCA80AB4863C55 CRC64;

Query Match 30.9%; Score 60; DB 1; Length 888;
Best Local Similarity 63.2%; Pred. No. 30;
Matches 12; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 4 PTLROWLAARAGGGGGG 22
DB 196 POWRW-----AGGGGGGG 210

RESULT 22
SUS_DROME
ID SUS_DROME STANDARD; PRT; 1322 AA.
AC P22293;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Suppressor of sable protein.
GN SU(S).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=91117256; PubMed=1703632;
RA Voelker R.A., Gibson W., Graves J.P., Sterling J.F., Eisenberg M.T.;
RT "The Drosophila suppressor of sable gene encodes a polypeptide with
RL regions similar to those of RNA-binding proteins.";
RN Mol. Cell. Biol. 11:894-905(1991).
RP [2]
RP SEQUENCE OF 1-9 FROM N.A.
RX MEDLINE=911169252; PubMed=1963868;
RA Voelker R.A., Graves J.P., Gibson W., Eisenberg M.T.;
RT "Mobile element insertions causing mutations in the Drosophila
RP suppressor of sable locus occur in DNase I hypersensitive subregions
of 5'-transcribed nontranslated sequences.";
-!- FUNCTION: AFFECTS THE TRANSCRIPT LEVELS OF THOSE ALLELES THAT IT
SUPPRESSES. MAY BE INVOLVED IN RNA METABOLISM.
-!- SUBCELLULAR LOCATION: Nuclear.
-!- DEVELOPMENTAL STAGE: AT ALL STAGES.
-!- SIMILARITY: HAS REGIONS SIMILAR TO THOSE OF RNA-BINDING PROTEINS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M57889; AAA28920.1;
EMBL; X59364; CAA42010.1;
PIR; A39612; A39612.
FlyBase; FBgn0003575; su(s).
InterPro; IPR000571; zf-CCCH.
Pfam; PF00642; zf-CCCH; 2.
RNA-binding; Nuclear protein.
FT DOMAIN 138 327
FT DOMAIN 446 474
FT DOMAIN 1087 1162
FT SEQUENCE 1322 AA; 143555 MW; D5F534E95702EA08 CRC64;

Query Match 30.9%; Score 60; DB 1; Length 1322;
Best Local Similarity 68.8%; Pred. No. 43;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPTLRQ 30
DB 1159 GGGGGGGVLPNLSQ 1174

RESULT 23
SOX1_MOUSE
ID SOX1_MOUSE STANDARD; PRT; 391 AA.
AC P53783;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SOX-1 protein.
GN SOX1 OR SOX-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=96189340; PubMed=8625802;
RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,
RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,
RA Lovell-Badge R.;
RT "A comparison of the properties of Sox-3 with Sry and two related
genes, Sox-1 and Sox-2.";
RN Development 122:509-520(1996).
-!- SUBCELLULAR LOCATION: Nuclear (Probable).
-!- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.
-!- SIMILARITY: CONTAINS 1 HMG BOX.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

Query Match
30.48; Score 59; DB 1; Length 401;

Best Local Similarity 64.7%; Pred. No. 19;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGIEGPT 27
Db 37 ASGTGGGGGGGASGPT 53
| : ||||| | |

RESULT 26
ONC2_HUMAN STANDARD; PRT; 485 AA.
ID OC95948;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE One cut domain family member 2 (ONECUT-2 transcription factor) (OC-2).
GN OCUCUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99115605; PubMed=9915796;
RA Jacquemin P., Lannoy V., Rousseau G.G., Lemaigre F.P.;
RT "OC-2, a novel mammalian member of the ONECUT class of homeodomain
transcription factors whose function in liver partially overlaps with
that of hepatocyte nuclear factor-6";
RL J. Biol. Chem. 274:2665-2671(1999).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES THE TRANSCRIPTION
OF A NUMBER OF LIVER GENES SUCH AS HNF3B.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 CUT DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEODOMAIN PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; Y18198; CAB38253.1; -
DR TRANSFAC; T03259; -
DR MIM; 604894; -
DR InterPro; IPR001350; CUT.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF02376; CUT; 1.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Activator.
FT DNA_BIND 305 391 CUT.
FT DNA_BIND 407 466 HOMEBOX.
FT DOMAIN 18 37 POLY-GLY.
FT DOMAIN 62 66 POLY-PRO.
FT DOMAIN 75 82 POLY-ALA.
FT DOMAIN 152 165 POLY-HIS.
FT DOMAIN 298 303 POLY-SER.
SQ SEQUENCE 485 AA; 52482 MW; AF21E052EFBE5DA1 CRC64;

Query Match 30.4%; Score 59; DB 1; Length 485;
Best Local Similarity 65.0%; Pred. No. 22;
Matches 13; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPTLRQWLAA 34
Db 25 GGGGGGGGGGPGHQELLA 44
- ||||| | | |

RESULT 27
ZIN_HUMAN STANDARD; PRT; 753 AA.
ID Q9NRL3;
AC 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Zinedin.
GN ZIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347911; PubMed=10748158;
RA Castets F., Rakitina T., Gaillard S., Moqrach A., Mattei M.-G.,
RA Monneron A.;
RT "Zinedin, SG2NA, and striatin are calmodulin-binding, WD repeat
proteins principally expressed in the brain.";
RL J. Biol. Chem. 275:19970-19977(2000).
RN [2]
RP SEQUENCE OF 402-753 FROM N.A.
RX TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS CALMODULIN IN A CALCIUM DEPENDENT MANNER. MAY
FUNCTION AS SCAFFOLDING OR SIGNALING PROTEIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-BOUND (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE STRIATIN FAMILY OF WD-REPEAT PROTEINS.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -!- CAUTION: The name "Zinedin" probably originates from the name of
the famous soccer player from Marseille (Zinedine Zidane)!

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF212940; AAF29527.1; -
DR EMBL; BC004910; AAH04910.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 4.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Calmodulin-binding; Repeat; WD repeat; Coiled coil.
KW Calmodulin-binding; Repeat; WD repeat; Coiled coil (POTENTIAL).
FT DOMAIN 69 136 CALMODULIN-BINDING (POTENTIAL).
FT DOMAIN 165 182 CALMODULIN-BINDING (POTENTIAL).
FT REPEAT 436 475 WD 1.
FT REPEAT 489 528 WD 2.
FT REPEAT 542 581 WD 3.
FT REPEAT 587 628 WD 4.
FT REPEAT 635 674 WD 5.
FT REPEAT 677 716 WD 6.
FT REPEAT 723 752 WD 7.
FT SITE 71 79 CAVEOLIN-BINDING (POTENTIAL).
FT DOMAIN 6 14 POLY-ALA.
FT CONFLICT 402 404 LAD -> GTR (IN REF. 2).
SQ SEQUENCE 753 AA; 80581 MW; 4DA016A8FF7EDB5E CRC64;

Query Match 30.4%; Score 59; DB 1; Length 753;
Best Local Similarity 78.6%; Pred. No. 33;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 AGGGGGGGGIEGPT 27
| | | | | | | | | |

```

Db 44 AKGGGGGSGPGT 57
RESULT 28
ID ECR_LUCCU STANDARD; PRT; 757 AA.
AC O18531;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-ecdysone
  receptor) (20E receptor).
GN ECR OR NR1H1.
OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Oestroidea; Calliphoridae; Lucilia.
OX NCBI_TaxID=7375;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97449774; PubMed=9304790;
RA Hannan G.N., Hill R.J.;
RT "Cloning and characterization of LcEcR: a functional ecdysone
  receptor from the sheep blowfly Lucilia cuprina."
RL Insect Biochem. Mol. Biol. 27:479-488(1997).
CC -1- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE
  ELEMENTS (ECRES) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC -1- SIMILARITY: BELONGS TO THE NR1 SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75355; AAB81130.1; -
DR HSP; P20393; IAGY.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Strdhormone_receptor.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STRDHOFINGER.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
  Zinc-finger.
FT DOMAIN 1 300 MODULATING (POTENTIAL).
FT DNA_BIND 301 366 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 301 321 C4-TYPE.
FT ZN_FING 337 361 C4-TYPE.
FT DOMAIN 454 674 HORMONE-BINDING (POTENTIAL).
SQ SEQUENCE 757 AA; 83075 MW; C1511452ED37D359 CRC64;
Query Match 30.4%; Score 59; DB 1; Length 757;
Best Local Similarity 76.9%; Pred. No. 33;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 15 GGGGGGGGIGEGPT 27
  |||||:|
Db 129 GCGGGGVPGMT 141
RESULT 29
ID YHA_CHLRE STANDARD; PRT; 4499 AA.
ID DYHA_CHLRE
AC Q39610;
01-NOV-1997 (Rel. 35, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Dynein alpha chain, flagellar outer arm (DHC alpha).
GN ODA11 OR ODA-11.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS.
RX STRAIN=21GR;
RX MEDLINE=97329535; PubMed=9186009;
RA Mitchell D.R., Brown K.S.;
RT "Sequence analysis of the Chlamydomonas reinhardtii flagellar alpha
  dynein gene."
RL Cell Motil. Cytoskeleton 37:120-126(1997).
RN [2]
RP SEQUENCE OF 1142-4499 FROM N.A.
RX STRAIN=21GR;
RX MEDLINE=94274778; PubMed=8006077;
RA Mitchell D.R., Brown K.S.;
RT "Sequence analysis of the Chlamydomonas alpha and beta dynein heavy
  chain genes."
RL J. Cell Sci. 107:635-644(1994).
CC -1- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
  FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC -1- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND
  GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L26049; AAA57316.2; -
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001298; Filamin.
DR InterPro; IPR002909; IPT_TIG.
DR InterPro; IPR001798; Kelch.
DR InterPro; IPR001736; PLD.
DR Pfam; PF00630; Filamin; 1.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SM00382; AAA; 3.
DR SMART; SM00429; IPT; 1.
DR PROSITE; PS50194; FILAMIN_REPEAT; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
  Coiled coil.
FT REPEAT 425 534 FILAMIN.
FT DOMAIN 1261 1334 COILED COIL (POTENTIAL).
FT DOMAIN 1382 1450 COILED COIL (POTENTIAL).
FT DOMAIN 1836 1864 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 2655 2688 COILED COIL (POTENTIAL).
FT DOMAIN 3003 3023 COILED COIL (POTENTIAL).
FT DOMAIN 3170 3262 COILED COIL (POTENTIAL).
FT DOMAIN 3486 3515 COILED COIL (POTENTIAL).
FT NP_BIND 1716 1723 ATP (POTENTIAL).
FT NP_BIND 2019 2026 ATP (POTENTIAL).
FT NP_BIND 2369 2376 ATP (POTENTIAL).
FT NP_BIND 2717 2754 ATP (POTENTIAL).
SQ SEQUENCE 4499 AA; 503606 MW; 319AC7FD30F1591A CRC64;
Query Match 30.4%; Score 59; DB 1; Length 4499;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 3 GPTLQWLAAAGGGGGG 22
  |||:|

```


Db 4194 GETLFTVVEVAGGGGGG 4213

RESULT 30

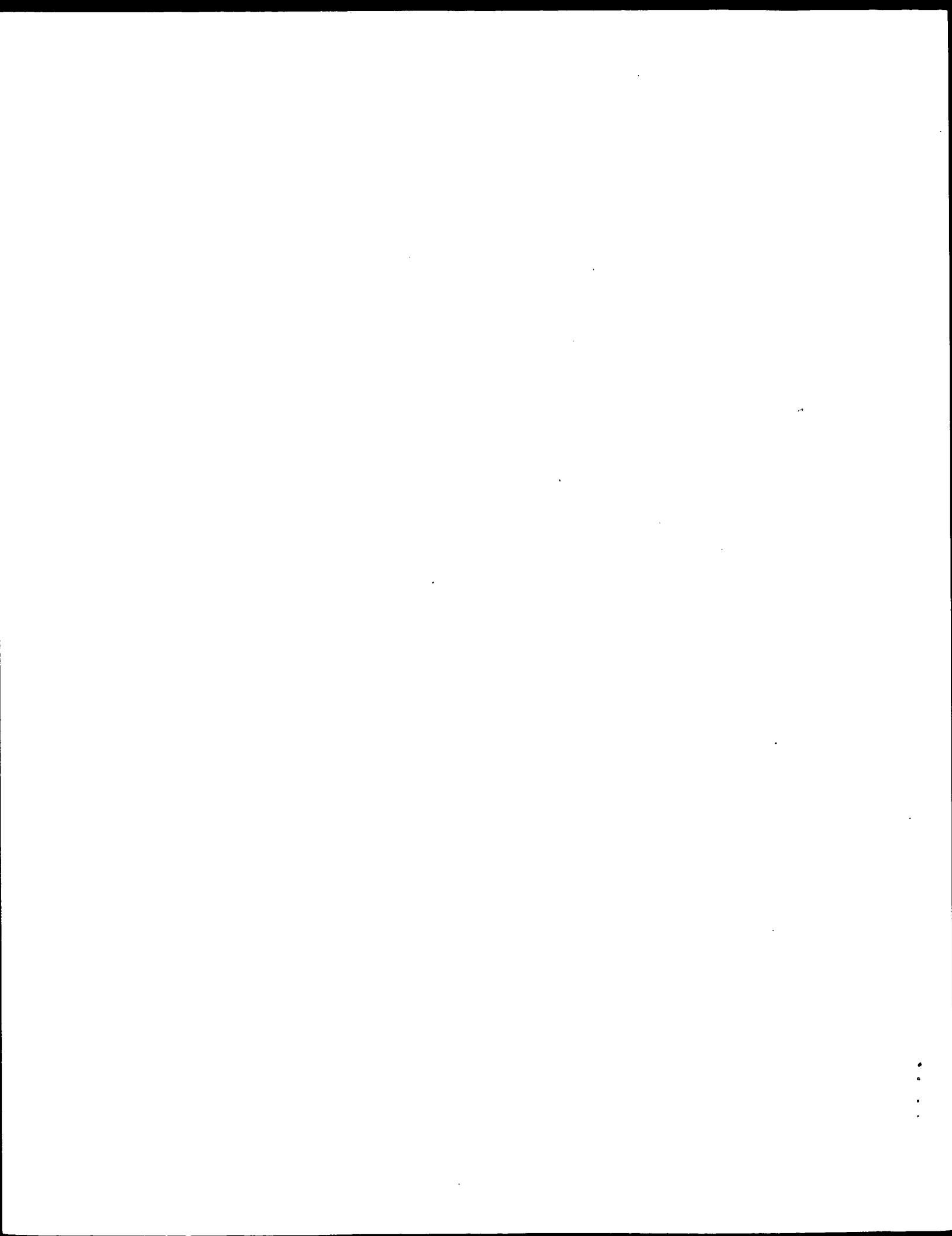
HXD9_HUMAN
ID HXD9_HUMAN STANDARD; PRT; 342 AA.
AC P28336;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Homeobox protein Hox-D9 (Hox-4C) (Hox-5.2).
GN HOXD9 OR HOX4C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spinal cord;
RX MEDLINE=92097538; PubMed=1756725;
RA Zappavigna V., Renucci A., Izpisua-Belmonte J.-C., Uriar G.,
RA Peschle C., Duboule D.;
RT "HOX4 genes encode transcription factors with potential auto- and
RT cross-regulatory capacities";
RL EMBO J. 10:4177-4187(1991).
RN [2]
RP SEQUENCE OF 264-342 FROM N.A.
RX MEDLINE=89306602; PubMed=2568311;
RA Oliver G., Sidell N., Fiske N., Heinzmann C., Mohandas T.,
RA Sparkes R.S., de Robertis E.M.;
RT "Complementary homeo protein gradients in developing limb buds";
RL Genes Dev. 3:641-650(1989).
RN [3]
RP SEQUENCE OF 275-340 FROM N.A.
RX MEDLINE=90098876; PubMed=2574852;
RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT "The human HOX gene family";
RL Nucleic Acids Res. 17:10385-10402(1989).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING LIMB BUDS.
CC -!- SIMILARITY: BELONGS TO THE ABD-B FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59372; CAA42016.1; -;
DR EMBL; X15506; CAA33528.1; -;
DR PIR; S18649; S18649.
DR PIR; S05958; S05958.
DR PIR; A32830; A32830.
DR HSP; P02834; I881.
DR TRANSFAC; T01424; -;
DR MIM; 142982; -;
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
Transcription regulation.
FT DOMAIN 115 149 GLY-RICH.
FT DOMAIN 121 130 POLY-GLY.
FT DOMAIN 165 178 SER/THR-RICH.

FT DNA_BIND 275 334 HOMEBOX.
FT CONFLICT 266 E -> A (IN REF. 2).
SQ SEQUENCE 342 AA; 35580 MW; 731981FE25C5ACD7 CRC64;

Query Match 30.2%; Score 58.5; DB 1; Length 342;
Best Local Similarity 44.8%; Pred. No. 19;
Matches 13; Conservative 2; Mismatches 9; Indels 5; Gaps 1;

QY 3 GPTLRQLAARAG-----GGGGGGGIEGP 26
Db 103 GRYVRSWMEPLPGFPGGAGGGGGGGG 131
| : | : | : ||||| ||
| : | : | : ||||| ||

Search completed: October 9, 2002, 09:00:10
Job time : 6.3831 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 12.8993 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-24
Perfect score: 194
Sequence: 1 IEPTPLRQLAARAGGGGGGIEPTPLRQLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	74	38.1	360 10 Q9LGC9	Q9Lgc9 oryza sativ
2	73.5	37.9	431 13 Q9PVG9	Q9pvg9 coturnix co
3	71	36.6	253 10 Q943K0	Q943k0 oryza sativ
4	70	36.1	439 10 Q9SDK6	Q9sdk6 oryza sativ
5	69	35.6	500 5 Q19476	Q19476 caenorhabdi
6	68.5	35.3	488 16 Q9CCCO	Q9ccco mycobacteri
7	68.5	35.3	518 2 Q49843	Q49843 mycobacteri
8	68	35.1	125 10 Q9LWC8	Q9lwc8 oryza sativ
9	68	35.1	776 3 Q9HEA4	Q9hea4 neurospora
10	67	34.5	170 5 Q9W033	Q9w033 drosophila
11	66.5	34.3	202 10 Q9FTZ5	Q9ftz5 oryza sativ
12	66.5	34.3	495 16 Q33230	Q33230 mycobacteri
13	66.5	34.3	496 2 Q9AD76	Q9ad76 streptomyc
14	66	34.0	377 13 Q9YHD0	Q9yhd0 streptomyc
15	66	34.0	529 10 Q9ASE5	Q9ase5 oryza sativ
16	66	34.0	612 4 Q9P270	Q9p270 homo sapien

17	65.5	33.8	243	10	Q9AR44	Q9ar44 oryza sativ
18	65.5	33.8	1548	4	Q9NYW9	Q9nyw9 homo sapien
19	65.5	33.8	2161	4	Q9Y566	Q9y566 homo sapien
20	65	33.5	447	13	O73628	O73628 anolis caro
21	65	33.5	452	5	Q9VJK4	Q9vjk4 drosophila
22	64	33.0	309	5	Q9VW01	Q9vv01 drosophila
23	64	33.0	331	5	Q9U211	Q9u211 caenorhabdi
24	64	33.0	333	5	Q9U210	Q9u210 caenorhabdi
25	64	33.0	422	5	O96755	O96755 branchiosto
26	63.5	32.7	207	10	Q94IW9	Q94iw9 oryza sativ
27	63.5	32.7	584	10	Q9L116	Q9l116 oryza sativ
28	63	32.5	66	12	Q91BC5	Q91bc5 spodoptera
29	63	32.5	137	10	Q9M6A1	Q9m6a1 catharanthu
30	63	32.5	160	10	Q9M699	Q9m699 catharanthu
31	63	32.5	186	10	Q942R8	Q942r8 oryza sativ
32	63	32.5	474	4	Q96SQ2	Q96sq2 homo sapien
33	63	32.5	490	10	O04270	O04270 chlamydomon
34	63	32.5	688	4	Q9BYD8	Q9byd8 homo sapien
35	63	32.5	689	4	Q96JG7	Q96jg7 homo sapien
36	63	32.5	752	4	O96L34	O96l34 homo sapien
37	63	32.5	841	10	O9SXI9	O9sxi9 oryza sativ
38	62.5	32.2	775	4	Q9C011	Q9c011 homo sapien
39	62	32.0	165	2	Q9AF15	Q9af15 mycobacteri
40	62	32.0	286	13	Q9PUX6	Q9pux6 gadus morhu
41	62	32.0	334	11	Q9JKB4	Q9jkb4 mus musculu
42	62	32.0	381	10	Q9LD54	Q9ld54 oryza sativ
43	62	32.0	540	2	Q93H33	Q93h33 streptomyc
44	62	32.0	642	13	Q9PUD8	Q9pud8 lampetra fl
45	62	32.0	664	5	Q9NEC7	Q9nec7 leishmania

ALIGNMENTS

RESULT 1

Q9LGC9 ID Q9LGC9 PRELIMINARY; PRT; 360 AA.
AC Q9LGC9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE PUTATIVE ZINC FINGER PROTEIN.
GN P0462H08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0462H08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002525; BAB07996.1; -
DR InterPro: IPR000571; Zf-CCCH.
DR Pfam: PF00684; Zf-CCCH; 4.
DR SMART: SM00356; Znf_C3H1; 4.
SQ SEQUENCE 360 AA; 37368 MW; 5105598D7E1C77B2 CRC64;

Query Match 38.1%; Score 74; DB 10; Length 360;
Best Local Similarity 56.0%; Pred. No. 1.4;
Matches 14; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 IEPTPLRQLAARAGGGGGGIEPTPLRQLAARA 36

Db 26 LEGPWRMGLGGGGGGGGGGGGGG 50

RESULT 2

Q9PVG9 ID Q9PVG9 PRELIMINARY; PRT; 431 AA.

RESULT 6

Q9CCCO PRELIMINARY; PRT; 488 AA.
 AC Q9CCCO;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE POSSIBLE ATP/GTP-BINDING PROTEIN.
 GN ML0997.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL583920; CAC31378.1; -.
 DR Leproma; ML0997; -.
 DR InterPro; IPR000765; GTP1_OBG.
 DR PRINTS; PR00326; GTP1OBG.
 KW Complete proteome.
 SQ SEQUENCE 488 AA; 52800 MW; 188918856F9774AA CRC64;

Query Match 35.3%; Score 68.5; DB 16; Length 488;
 Best Local Similarity 46.7%; Pred. No. 7.4;
 Matches 14; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGGGGGIEGP 26

Db 189 PRLRGESMSRQVGGGAGGGGGLRGP 218

RESULT 7

Q49843 PRELIMINARY; PRT; 518 AA.
 AC Q49843;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HFLX.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith D.R.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robison K.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U00019; AAA17274.1; -.
 SQ SEQUENCE 518 AA; 56001 MW; 6641916CC84F374B CRC64;

Query Match 35.3%; Score 68.5; DB 2; Length 518;
 Best Local Similarity 46.7%; Pred. No. 7.8;
 Matches 14; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGGGGGIEGP 26
 Db 219 PRLRGESMSRQVGGGAGGGGGLRGP 248

RESULT 8

Q9LWC8 PRELIMINARY; PRT; 125 AA.
 AC Q9LWC8;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0483F08.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002094; BAA96216.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 125 AA; 13396 MW; C609D8D0B07BC505 CRC64;

Query Match 35.1%; Score 68; DB 10; Length 125;
 Best Local Similarity 42.9%; Pred. No. 2.2;
 Matches 18; Conservative 2; Mismatches 8; Indels 14; Gaps 2;

QY 2 EGPTLRQWLARA-----GGGGGGGIEGPTLRQ 30

Db 83 EGAAAR-WRAARSPARGGRRRRGGGGGGRPRRRR 123

RESULT 9

Q9HEA4 PRELIMINARY; PRT; 776 AA.
 AC Q9HEA4;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN BLIA5.200.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL451109; CAC18624.2; -.
 KW Hypothetical protein.
 SQ SEQUENCE 776 AA; 82771 MW; C9BEA870D94A37DE CRC64;

Query Match 35.1%; Score 68; DB 3; Length 776;
 Best Local Similarity 57.7%; Pred. No. 13;
 Matches 15; Conservative 3; Mismatches 4; Indels 4; Gaps 2;

QY 15 GGGGGGGI---EG-PTLRQWLARA 36

Db 678 GGGGGGGVVDDDDGPDPAAGLAAQA 703

3.

QW	DNA-binding; Homeobox; Nuclear protein.
QW	SEQUENCE 377 AA; 37998 MW; C2DBC19402D3A172 CRC64;
Query Match	34.0%; Score 66; DB 13; Length 377;
Best Local Similarity	48.1%; Pred. No. 11;
Matches 13; Conservative	2; Mismatches 12; Indels 0; Gap
QY	2 EGPTLRQWLAAARAGGGGGGIEGPTL 28
Db	: : : : :
Db	265 QGYTAASYGVCEGGGGGGGGGPGYL 291
QY	QYASE5
ID	QYASE5 PRELIMINARY; PRT; 529 AA.
AC	QYASE5;
DT	01-JUN-2001 (TREMBLrel. 17, Created)
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE	P0456F08.14 PROTEIN.
GN	P0456F08.14.
OS	Oryza sativa (Rice).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC	Ehrhartoideae; Oryzaceae; Oryza.
NCBI_TaxID=4530;	
QY	SEQUENCE FROM N.A.
RP	STRAIN-CV. NIPPONBARE;
RC	Sasaki T., Matsumoto T., Yamamoto K.;
RT	"Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT	clone:P0456F08.14;
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF002901; BAB39414.1; -
DR	InterPro; IPR002937; Amino_oxidase.
DR	InterPro; IPR002020; NAD-binding.
DR	Pfam; PF01593; Amino_oxidase; 1.
QY	SEQUENCE 529 AA; 55981 MW; 0A5DA5CDD076D24 CRC64;
Query Match	34.0%; Score 66; DB 10; Length 529;
Best Local Similarity	68.4%; Pred. No. 15;
Matches 13; Conservative	2; Mismatches 4; Indels 0; Gap
QY	6 LRQWLAAARAGGGGGGGGIE 24
Db	: : : : :
Db	151 LRAYQAARSAGGGGGGKE 169
QY	QY270
ID	QY270 PRELIMINARY; PRT; 612 AA.
AC	QY270;
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE	K1AAL458 PROTEIN (FRAGMENT).
GN	K1AAL458.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;	
QY	SEQUENCE FROM N.A.
RP	MEDLINE=20277482; PubMed=10819331;
RX	Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT	"Prediction of the coding sequences of unidentified human
RT	genes.XVII. The complete sequences of 100 new cDNA clones from brain
RT	which code for large proteins in vitro.";
RL	DNA Res. 7:143-150(2000).
DR	EMBL; AB040891; BAA95982.1; -
FT	NON_TER 1
QY	SEQUENCE 612 AA; 65593 MW; 9AA4061D21E1E9FD CRC64;

OS Homo sapiens (Human),
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cusley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Mizay D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003528; AAF49521.1; -
 DR FlyBase: FBgn0036583; CG13055.
 SQ SEQUENCE 309 AA; 33224 MW; 9DAEB67784852A93 CRC64;

Query Match 33.0%; Score 64; DB 5; Length 309;
 Best Local Similarity 57.9%; Pred. No. 14;
 Matches 11; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 12 ARAGGGGGGGGIEGPTLRQ 30

Db 94 SRSGGGGGGGAGVTLQE 112

RESULT 23

Q90211 ID Q90211 PRELIMINARY; PRT; 331 AA.
 AC Q90211;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Y41C4A.4A PROTEIN.
 GN Y41C4A.4A.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Steward C.A.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C.elegans*: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
 DR EMBL: AL032627; CAB54381.1; -
 DR InterPro: IPR001871; bZIP.
 DR InterPro: IPR003102; pKID.

DR Pfam: PF00170; bZIP; 1.
 DR Pfam: PF02173; pKID; 1.
 DR SMART: SM00338; BRLZ; 1.
 DR PROSITE: PS00036; BZIP_BASIC; 1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 331 AA; 34985 MW; A414C19D4ADCC91E CRC64;

Query Match 33.0%; Score 64; DB 5; Length 331;
 Best Local Similarity 76.9%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27

Db 167 GGGGGGGGVPGPS 179

RESULT 24

Q90210 ID Q90210 PRELIMINARY; PRT; 333 AA.
 AC Q90210;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Y41C4A.4B PROTEIN.
 GN Y41C4A.4B.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Steward C.A.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C.elegans*: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
 DR EMBL: AL032627; CAB54382.1; -
 DR InterPro: IPR001871; bZIP.
 DR InterPro: IPR003102; pKID.
 DR Pfam: PF00170; bZIP; 1.
 DR Pfam: PF02173; pKID; 1.
 DR SMART: SM00338; BRLZ; 1.
 DR PROSITE: PS00036; BZIP_BASIC; 1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 333 AA; 35261 MW; BF02CE6398F6D058 CRC64;

Query Match 33.0%; Score 64; DB 5; Length 333;
 Best Local Similarity 76.9%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27

Db 169 GGGGGGGGVPGPS 181

RESULT 25

O96755 ID O96755 PRELIMINARY; PRT; 422 AA.
 AC O96755;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE INTERMEDIATE FILAMENT PROTEIN E1.
 OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7740;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=CV. NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT  clone:P0708G02.";
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AP001539; BAA92923.1; -
DR  HSSP; P00950; 5PGM.
DR  InterPro; IPR001345; PG_mutase.
DR  Pfam; PF00300; PGAM; 1.
DR  PRINTS; PR01248; TYPEKERATIN.
SQ  SEQUENCE 422 AA; 44892 MW; 85FE742F07751B24 CRC64;

Query Match 33.08; Score 64; DB 5; Length 422;
Best Local Similarity 61.98; Pred. No. 19;
Matches 13; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

QY 15 GGGGGGGGIEG-----PTLR 29
    |||||
Db 92 GGGGGGGGSGMWTKEPTMR 112
    |||||

RESULT 26
Q941W9
ID Q941W9 PRELIMINARY; PRT; 207 AA.
AC Q941W9;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE P0037C04.13 PROTEIN.
GN P0037C04.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0037C04.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003233; BAB55526.1; -
SQ SEQUENCE 207 AA; 21266 MW; F514ABC36A6DC403 CRC64;

Query Match 32.78; Score 63.5; DB 10; Length 207;
Best Local Similarity 45.58; Pred. No. 11;
Matches 15; Conservative 4; Mismatches 5; Indels 9; Gaps 2;

QY 11 AARAGGGG-----GGGIEGPTLRQWLAARA 36
    |||||
Db 122 AVAGGGGGCSDAVQAGG--GGAVQWCASES 152
    |||||

RESULT 27
Q941L16
ID Q941L16 PRELIMINARY; PRT; 584 AA.
AC Q941L16;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=CV. NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT  clone:P0708G02.";
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AP001539; BAA92923.1; -
DR  HSSP; P00950; 5PGM.
DR  InterPro; IPR001345; PG_mutase.
DR  Pfam; PF00300; PGAM; 1.
DR  PRINTS; PR01248; TYPEKERATIN.
SQ  SEQUENCE 584 AA; 63515 MW; 351C684C8BBB99CF CRC64;

Query Match 32.78; Score 63.5; DB 10; Length 584;
Best Local Similarity 48.38; Pred. No. 30;
Matches 14; Conservative 2; Mismatches 8; Indels 5; Gaps 1;

QY 7 ROWLAARA-----GGGGGGGIEGPTLRQ 30
    |||||
Db 113 RWTATRSSDPGIGSGGGGGGEGAPTRRR 141
    |||||

RESULT 28
Q91BC5
ID Q91BC5 PRELIMINARY; PRT; 66 AA.
AC Q91BC5;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 7.0 KDA PROTEIN.
OS Spodoptera litura nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46242;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang H.;
RA Pang Y., Yu J., Wang L., Hu X., Bao W., Li G., Chen C., Han H., Hu S.,
RA Yang H.;
RT "Sequence Analysis of the Spodoptera litura Multicapsid
RT Nucleopolyhedrovirus Genome.";
RL Virology 287:391-404(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Yu J., Wang L., Hu X., Pang Y.;
RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF325155; AAL01786.1; -
RW Hypothetical protein.
SQ SEQUENCE 66 AA; 6998 MW; C5G26A8FFA9C9E7C CRC64;

Query Match 32.58; Score 63; DB 12; Length 66;
Best Local Similarity 68.88; Pred. No. 3.9;
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 13 RAGGGGGGGGIEGPTL 28
    :|||||
Db 19 RSGGGGGGGGVVGAML 34
    :|||||

RESULT 29
Q9M6A1
ID Q9M6A1 PRELIMINARY; PRT; 137 AA.
AC Q9M6A1;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PUTATIVE GLYCINE-RICH RNA BINDING PROTEIN 1.
OS GRP-1.
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 OC Vincaeae; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Clastre M., Chenieux J.C., Rideau M., Hamdi S.;
 RT "Genes encoding glycine-rich Catharanthus roseus proteins with RNA-
 binding motifs."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF200321; AAF31402.1; -
 DR HSSP: P09651; 1HA1.
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; rrm; 1.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PS0102; RRM; 1.
 DR PROSITE: PS00030; RRM_RNP.1; 1.
 SQ SEQUENCE 137 AA; 14162 MW; 4FABAD9C7A989FC CRC64;

Query Match 32.5%; Score 63; DB 10; Length 137;
 Best Local Similarity 50.0%; Pred. No. 8.1;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 TLQWLAAAGGGGGGGIEGP 26
 I : : : : :
 Db 80 TVNEAQRSGGGGGGGFRGP 101

RESULT 30

O9M699
 ID O9M699 PRELIMINARY; PRT: 160 AA.
 AC O9M699;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE GLYCINE-RICH RNA-BINDING PROTEIN 2.
 GRP-2.
 OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 OC Vincaeae; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Courtois M., Chenieux J.-C., Hamdi S., Rideau M.,
 RA Clastre M.;
 RT "Cloning of two cDNAs encoding crGRP2 and crGRP3 (Accession Nos.
 RT AF200323 and AF200322), the first members of the RRM-GRP family in
 RT Catharanthus roseus (PGR00-049).";
 RL Plant Physiol. 122:1459-1459(2000).
 DR EMBL: AF200323; AAF31404.1; -
 DR HSSP: P09651; 1HA1.
 DR InterPro: IPR002952; Eggshell.
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; rrm; 1.
 DR PRINTS: PR01228; EGGSHLL.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PS0102; RRM; 1.
 DR PROSITE: PS00030; RRM_RNP.1; 1.
 SQ SEQUENCE 160 AA; 16264 MW; DCDC9F63C983F5F2 CRC64;

Query Match 32.5%; Score 63; DB 10; Length 160;
 Best Local Similarity 50.0%; Pred. No. 9.4;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 TLQWLAAAGGGGGGGIEGP 26
 I : : : : :
 Db 80 TVNEAQRSGGGGGGGFRGP 101

Search completed: October 9, 2002, 09:03:06
 Job time : 15.9826 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 15.2881 Seconds

(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-25

Perfect score: 183
Sequence: 1 GGIEGPTLRQMLARAGNGIEGPTLRQMLARA 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```
1: A_Geneseq_032802:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	183	100.0	34 21 AAY96527	Thrombopoietin mim
2	171	93.4	32 21 AAB17297	TPO-mimetic peptid
3	171	93.4	32 21 AAY96520	Thrombopoietin mim
4	156	85.2	32 21 AAB17289	TPO-mimetic peptid
5	156	85.2	41 21 AAY96528	Thrombopoietin mim
6	156	85.2	42 21 AAB17281	TPO-mimetic peptid
7	156	85.2	42 21 AAB17308	Synthetic TWP-TMP
8	156	85.2	42 21 AAY96530	Thrombopoietin mim
9	156	85.2	269 21 AAY96531	Human IgG1 Fc TMP
10	152	83.1	268 21 AAB16959	Fc-TMP protein
11	147.5	80.6	31 21 AAB17288	TPO-mimetic peptid

12	147	80.3	30 21 AAB17287	TPO-mimetic peptid
13	145.5	79.5	33 21 AAB17290	TPO-mimetic peptid
14	145	79.2	34 21 AAB17291	TPO-mimetic peptid
15	145	79.2	36 21 AAB17306	TPO-mimetic peptid
16	145	79.2	35 21 AAY96526	Thrombopoietin mim
17	144.5	79.0	35 21 AAB17292	TPO-mimetic peptid
18	144	78.7	36 21 AAB16963	TPO-mimetic peptid
19	144	78.7	36 21 AAB17293	TPO-mimetic peptid
20	144	78.7	36 21 AAB17301	TPO-mimetic peptid
21	144	78.7	36 21 AAB17303	TPO-mimetic peptid
22	144	78.7	36 21 AAB17307	TPO-mimetic peptid
23	144	78.7	36 21 AAY96523	Thrombopoietin mim
24	144	78.7	36 21 AAY96524	Thrombopoietin mim
25	144	78.7	36 21 AAY96525	Synthetic TWP-TMP
26	144	78.7	42 21 AAB17282	TWP-TMP-Fc protein
27	144	78.7	60 21 AAB17311	TPO-mimetic peptid
28	144	78.7	269 21 AAB17294	TPO-mimetic peptid
29	143.5	78.4	37 21 AAB17295	TPO-mimetic peptid
30	143	78.1	38 21 AAB17304	TPO-mimetic peptid
31	142.5	77.9	39 21 AAB17305	TPO-mimetic peptid
32	142.5	77.9	39 21 AAB17302	TPO-mimetic peptid
33	142	77.6	40 21 AAB17296	TPO-mimetic peptid
34	141	77.0	42 21 AAB17285	TPO-mimetic peptid
35	140.5	76.8	29 21 AAB17286	TPO-mimetic peptid
36	134	73.2	28 21 AAB16970	TPO-mimetic peptid
37	133.5	73.0	29 21 AAB16973	TPO-mimetic peptid
38	133.5	73.0	31 21 AAB16974	TPO-mimetic peptid
39	133.5	73.0	31 21 AAB16975	TPO-mimetic peptid
40	127.5	69.7	29 21 AAB16976	TPO-mimetic peptid
41	120.5	65.8	29 21 AAB16977	TPO-mimetic peptid
42	120.5	65.8	29 21 AAB16978	TPO-mimetic peptid
43	118	64.5	36 21 AAB17298	TPO-mimetic peptid
44	118	64.5	36 21 AAB17299	TPO-mimetic peptid
45	118	64.5	36 21 AAY96521	Cyclic or linear t

ALIGNMENTS

RESULT 1	AA96527	standard; peptide; 34 AA.
ID	AA96527	standard; peptide; 34 AA.
XX	AA96527	
AC	AA96527	
XX	AA96527	
DT	04-SBP-2000	(first entry)
XX	04-SBP-2000	
DE	Thrombopoietin mimetic peptide compound 8.	
XX	Thrombopoietin mimetic; TWP; TPO; platelet; megakaryocyte; production;	
KW	anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;	
KW	immunosuppressive; anti-inflammatory; linker.	
XX	Synthetic.	
OS	Synthetic.	
XX	Synthetic.	
EH	Key	Location/Qualifiers
FT	Modified-site	1
FT	Peptide	/note= "optionally linked to an Fc molecule"
FT	Peptide	3..16
FT	Peptide	/label= TWP_1
FT	Peptide	17..20
FT	Peptide	/label= linker
FT	Peptide	21..34
PN	Peptide	/label= TWP_2
XX	W0200024770-A2.	
PD	04-MAY-2000.	
XX	22-OCT-1999;	99WO-US24834.
PF	23-OCT-1998;	98US-0105348.
XX	23-OCT-1998;	98US-0105348.
PR	23-OCT-1998;	98US-0105348.
XX	23-OCT-1998;	98US-0105348.

PA (AMGE-) AMGEN INC.
 XX
 PI Liu C, Feige U, Cheetham J;
 XX
 DR WPI; 2000-365108/31.
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 PS
 PS Claim 16; Page 64; 91pp; English.
 XX
 XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP-1-(L1)-TMP-2),
 CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
 CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N,
 CC L, F, S, T, K, H, or E; X-1-0 = L, I, V, A, F, M, or K; X-1-1 = A, I, V,
 CC T, V, N, O or G; X-1-2 = A, I, V, L, F, T, R, E, or G; L-1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 SQ Sequence 34 AA:
 Query Match 100.0%; Score 183; DB 21; Length 34;
 Best Local Similarity 100.0%; Pred. No. 1,3e-18;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GIEGPTLROMLAARAGNGIEGPTLROMLAARA 34
 DB 1 GIEGPTLROMLAARAGNGIEGPTLROMLAARA 34
 RESULT 2
 AAB17297 standard; Peptide; 32 AA.
 AC AAB17297;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:353.
 XX
 KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumor necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.
 KW
 XX Synthetic.
 OS
 OS WO200024782-A2.
 PN
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;

XX
 DR WPI; 2000-350702/30.
 XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 1; Page 320; 608pp; English.
 XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 32 AA:
 Query Match 93.4%; Score 171; DB 21; Length 32;
 Best Local Similarity 100.0%; Pred. No. 5.8e-17;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 IESEPTLROMLAARAGNGIEGPTLROMLAARA 34
 DB 1 IESEPTLROMLAARAGNGIEGPTLROMLAARA 32
 RESULT 3
 AAY96520 standard; peptide; 32 AA.
 AC AAY96520;
 XX
 DT 04-SEP-2000 (first entry)
 XX
 DE Thrombopoietin mimetic peptide compound 1.
 XX
 KW Thrombopoietin; mimetic; TMP; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-venetic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.
 KW
 XX Synthetic.
 OS
 OS Key Location/Qualifiers
 FH Modified-site 1 /note="optionally linked to an Fc molecule"
 FT Peptide 1..14 /label= TMP_1
 FT Peptide 15..18 /label= linker
 FT Peptide 19..32 /label= TMP_2
 FT Modified-site 32 /note="optionally linked to an Fc molecule"
 FT
 FT
 FT
 PN WO200024770-A2.
 PN
 XX
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24834.
 XX

PR 23-OCT-1998; 98US-0105348.
XX
XX (AMGE-) AMGEN INC.
XX
XX Liu C, Feige U, Cheetham J;
XX WPI; 2000-365108/31.
DR
XX Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thrombocytopenia
XX
XX
PS Claim 16; Page 61; 91pp; English.
XX
XX A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker (fTMP-1-(L1)-nTMP-2),
CC is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
CC X-4 = F; X-5 = T or S; X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N,
CC or E; X-9 = W, Y or F; X-10 = L, I, V, A, F, M, or K; X-11 = A, I, V,
CC L, F, S, T, K, H, or E; X-12 = A, I, V, L, F, G, S, or Q; X-13 = R, K,
CC T, V, N, Q or G; X-14 = A, I, V, L, F, T, R, E, or G; L-1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.
XX
XX Sequence 32 AA:
SQ
Query Match 93.4%; Score 171; DB 21; Length 32;
Best Local Similarity 100.0%; Pred. No. 5.8e-17;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DY 3 IEPTLRQWLAAARAGPGIEGPTLRQWLARA 34
DB 1 IEPTLRQWLAAARAGPGIEGPTLRQWLARA 32
RESULT 4
AAB17289
ID AAB17289 standard; Peptide; 32 AA.
XX
XX AAB17289;
XX
XX 31-OCT-2000 (first entry)
XX
XX TPO-mimetic peptide sequence SEQ ID NO:345.
XX
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNP; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
XX Synthetic.
XX
XX WO200024782-A2.
XX
XX 04-MAY-2000.
XX
XX 25-OCT-1999; 99WO-US25044.
XX
XX 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
XX (AMGE-) AMGEN INC.

XX
XX Feige U, Liu C, Cheetham J, Boone TC;
PI
XX WPI; 2000-350702/30.
DR
XX
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
XX Example 1; Page 316; 608pp; English.
XX
XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)-a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)-c-p1, -(L1)-c-p1-(L2)-d-p2,
CC -(L1)-c-p1-(L2)-d-p2-(L3)-e-p3, or -(L1)-c-p1-(L2)-d-p2-(L3)-e-p3-(L4)-f-p4
CC where p1, p2, p3, and p4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
XX Sequence 32 AA:
SQ
Query Match 85.2%; Score 156; DB 21; Length 32;
Best Local Similarity 93.8%; Pred. No. 7.1e-15;
Matches 30; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DY 3 IEPTLRQWLAAARAGPGIEGPTLRQWLARA 34
DB 1 IEPTLRQWLAAARAGPGIEGPTLRQWLARA 32
RESULT 5
AAV96528
ID AAV96528 standard; peptide; 41 AA.
XX
XX AAV96528;
XX
XX 04-SEP-2000 (first entry)
XX
XX Thrombopoietin mimetic peptide compound 9.
XX
XX Thrombopoietin mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1
FT Peptide 6..19
FT Peptide /note= "optionally linked to an Fc molecule"
FT Peptide /label= TMP_1
FT Peptide 20..27
FT Peptide /label= linker
FT Peptide 28..41
XX /label= TMP_2
XX
XX WO200024770-A2.
XX
XX 04-MAY-2000.
XX
XX 22-OCT-1999; 99WO-US24834.
XX

PR 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thrombocytopenia

XX Claim 16; Page 65; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker (TMP-1-(L1)-TMP-2),
CC is new, TMP-1 and TMP-2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X₁-X₁₀, X₂-X₁₁, X₂-X₁₂,
CC X₂-X₁₃, X₂-X₁₄, X₁-I, A, V, L, S or R; X₂-E, D, K or V; X₃-G or A;
CC X₁-X₁₄, X₁-I, A, V, L, S or R; X₂-E, D, K or V; X₃-G or A;
CC X₄-P; X₅-W, Y or F; X₆-L, I, V, A, F, M, or K; X₁₁-A, I, V,
CC or E; X₉-W, Y or F; X₁₀-L, I, V, A, F, M, or K; X₁₁-A, I, V,
CC L, F, S, T, K, H, or E; X₁₂-A, I, V, L, F, T, R, E, or G; L1 = linker
CC T, V, N, Q or G; X₁₄-A, I, V, L, F, T, R, E, or G; L1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 41 AA:

Query Match 85.2%; Score 156; DB 21; Length 41;

Best Local Similarity 84.2%; Pred. No. 9.4e-15; Mismatches 2; Indels 4; Gaps 1;

Matches 32; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

0Y 1 GGIEGPTLRQWLARA---GPNIGEGPTLRQWLARA 34
|||||
DB 4 GGIEGPTLRQWLARAAGGGGGGIEGPTLRQWLARA 41

RESULT 6

AAB17281

ID AAB17281 standard; Peptide: 42 AA.

XX AAB17281;

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:337.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antitumour; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CT1A4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

XX Synthetic.

OS WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

XX Disclosure; Page 313; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)-c-p1, -(L1)-c-p1-(L2)d-p2,
CC -(L1)-c-p1-(L2)d-p2-(L3)e-p3, or -(L1)-c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4
CC where p1, p2, p3, and p4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antitumour, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AA65443
CC to AA65526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 42 AA:

Query Match 85.2%; Score 156; DB 21; Length 42;

Best Local Similarity 84.2%; Pred. No. 9.6e-15; Mismatches 2; Indels 4; Gaps 1;

Matches 32; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

0Y 1 GGIEGPTLRQWLARA---GPNIGEGPTLRQWLARA 34
|||||
DB 5 GGIEGPTLRQWLARAAGGGGGGIEGPTLRQWLARA 42

RESULT 7

AAB17308

ID AAB17308 standard; Peptide: 42 AA.

XX AAB17308;

DT 31-OCT-2000 (first entry)

XX Synthetic TPO-TMP gene construction peptide SEQ ID NO:374.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antitumour; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CT1A4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

XX Homo sapiens.

OS Synthetic.

PN WO200024782-A2.

PD 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

Example 2A; Page 48; 91pp; English.

gene encoding a thrombopoietin mimetic peptide (TMP), which was then fused in-frame to the EC region of the human T-301

CC dimer joined by a

ery Match

Best Local Similarity

C
C
E
D
P
A
R
T
I
C
L
E

RESULT 9
A1V06531

XX
AC
XX
DT

AAY96531;
04-SEP-2000 (first entry)

XX	
KW	Immunoglobulin; IgG1; Fc; thrombo-
KW	megakaryocyte; production; anti-th
KW	anti-thrombotic; thrombotic; anti-th

AA Homo sapiens.
OS
XX
PN W0200024770-A2.
XX
04 MAR 2000
END

XX
XX
22-OCT 1966

XX	22-00cm-1000	00000	01000000
DB			

XX
BA (AMCE-) AMGEN TUC

XX	Fin C	Estad n	Chatham
PT			

XX
DR
WPT: 2000-365108/31

DR N-PSDB; AAA29229.

production of platelets or platelet precursors useful for thrombopoietic peptides which activate mpl receptors and increase the

XX diseases which involve thrombocytopenia

Example 2A; Page 49-50; 91pp; English.

CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP_1-(L1)_nTMP_2),
 CC is new, TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X2-X1.0, X2-X1.1, X2-X1.2,
 CC X2-X1.3, X2-X1.4, X1-X1.0, X1-X1.1, X1-X1.2, X1-X1.3, and
 CC X1-X1.4, X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
 CC X4 = P; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
 CC or E; X9 = W, Y or F; X1.0 = L, I, V, A, F, M, or K; X1.1 = A, I, V,
 CC L, F, S, T, K, H, or F; X1.2 = A, I, V, L, F, G, S, or Q; X1.3 = R, K,
 CC T, V, N, Q or G; X1.4 = A, I, V, L, F, T, R, E, or G; L1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 269 AA:

Query Match 85.2%; Score 156; DB 21; Length 269;
 Best Local Similarity 84.2%; Pred. No. 7.6e-14;
 Matches 32; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 GGIEGPTLRQWLARA---GPNIGEGPTLRQWLARA 34
 ||||||||||||||||| | |||||||||||||||||
 Db 232 GGIEGPTLRQWLARAAGGGGGGIEGPTLRQWLARA 269

RESULT 10

AAB16959
 ID AAB16959 standard; Protein: 268 AA.

XX AAB16959;

DT 31-OCT-2000 (first entry)

DE Fe-TMP-TMP protein sequence SEQ ID NO:8.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antilastmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

OS Homo sapiens.
 OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheatham J, Boone TC;

DR WPI: 2000-350702/30.

DR N-PSDB: AAA69445.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 182-183; 608pp; English.

XX The present invention describes composition of matter (I) comprising an

CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antilastmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 268 AA:

Query Match 83.1%; Score 152; DB 21; Length 268;
 Best Local Similarity 83.8%; Pred. No. 2.7e-13;
 Matches 31; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 1 GGIEGPTLRQWLARA---GPNIGEGPTLRQWLAR 33
 ||||||||||||||||| | |||||||||||||||||
 Db 232 GGIEGPTLRQWLARAAGGGGGGIEGPTLRQWLAR 268

RESULT 11

AAB17288
 ID AAB17288 standard; Peptide: 31 AA.

XX AAB17288;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:344.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antilastmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

OS Synthetic.
 OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheatham J, Boone TC;

DR WPI: 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 316; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is (X1)-a-P1-(X2)b, where: P1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)-c-P1, -(L1)c-P1-(L2)d-P2, -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3, or -(L1)-c-P1-(L2)d-P2-(L3)-e-P3-(L4)-f-P4

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is (X1)-a-p1-(X2)b, where: p1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)-c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4-(L1)-c-p1-(L2)d-p2-(L3)e-p3, or -(L1)-c-p1-(L2)d-p2-(L3)e-p3-(L4)f-p4 where p1, p2, p3, and p4 = are each independently sequences of pharmacologically active peptides L1, L2, L3, and L4 = are each

independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AA69443 to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Query Match 79.5%; Score 145.5; DB 21; Length 33;
Best Local Similarity 90.9%; Pred. No. 2.1e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 3 IECPTRLRWLAARA-GPNGIEGPTLRWLARA 34
1 IECPTRLRWLAARAAGGGGIEGPTLRWLARA 33

RESULT 14
AAB17291
ID AAB17291 standard; Peptide: 34 AA.
AC AAB17291;
XX 31-OCT-2000 (first entry)
XX TPO-mimetic peptide sequence SEQ ID NO:347.
DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.
OS Synthetic.
XX WO200024782-A2.
XX 04-MAY-2000.
XX 25-OCT-1999; 99WO-US25044.
XX 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX (AMGE-) AMGEN INC.
XX Feige U, Liu C, Cheetham J, Boone TC;
XX WPI: 2000-350702/30.
XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides; useful for treating cancer and
XX autoimmune diseases -
XX Example 1; Page 317; 608bp; English.
XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)-c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can

have cytostatic, antiasthmatic, thrombolytic and immunosuppressive 33
activities. DNAs, vectors and host cells from the present invention can 33
be used for producing pharmaceutical compositions. The compositions are 33
useful for treating cancer, asthma, thrombosis, or autoimmune diseases. 33
The use of an Fc domain (rather than a Fab domain) can provide a longer 33
half-life or incorporate functions such as Fc receptor binding, protein 33
A binding, complement fixation, and possibly placental transfer. AA69443 33
to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid 33
sequences used in the exemplification of the present invention. 33

Query Match 79.2%; Score 145; DB 21; Length 34;
Best Local Similarity 88.2%; Pred. No. 2.6e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

OY 3 IECPTRLRWLAARA-GPNGIEGPTLRWLARA 34
1 IECPTRLRWLAARAAGGGGIEGPTLRWLARA 34

RESULT 15
AAB17306
ID AAB17306 standard; Peptide: 36 AA.
AC AAB17306;
XX 31-OCT-2000 (first entry)
XX TPO-mimetic peptide sequence SEQ ID NO:362.
DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.
OS Synthetic.
XX WO200024782-A2.
XX 04-MAY-2000.
XX 25-OCT-1999; 99WO-US25044.
XX 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX (AMGE-) AMGEN INC.
XX Feige U, Liu C, Cheetham J, Boone TC;
XX WPI: 2000-350702/30.
XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides; useful for treating cancer and
XX autoimmune diseases -
XX Example 1; Page 324; 608bp; English.
XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)-c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
XX activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAB69443
CC to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

SO Sequence 35 AA:

Query Match 79.0%; Score 144.5; DB 21; Length 35;
Best Local Similarity 85.7%; Pred. No. 3.1e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 3 IEGPTLRQWLAAARA---GPNIEGPTLRQWLAAARA 34
|||||
Db 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 35

RESULT 18

AAB16963 ID AAB16963 standard; Protein; 36 AA.

AC AAB16963;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide TWP-TWP SEQ ID NO:14.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI: 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

PS Disclosure; Page 190; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
XX activities. DNAs, vectors and host cells from the present invention can
XX be used for producing pharmaceutical compositions. The compositions are
XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.

CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAB69443
CC to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

SO Sequence 36 AA:

Query Match 78.7%; Score 144; DB 21; Length 36;
Best Local Similarity 83.3%; Pred. No. 3.8e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 3 IEGPTLRQWLAAARA---GPNIEGPTLRQWLAAARA 34
|||||
Db 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 19

AAB17293 ID AAB17293 standard; Peptide; 36 AA.

AC AAB17293;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:349.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI: 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

PS Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
XX have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
XX activities. DNAs, vectors and host cells from the present invention can
XX be used for producing pharmaceutical compositions. The compositions are
XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
XX The use of an Fc domain (rather than a Fab domain) can provide a longer
XX half-life or incorporate functions such as Fc receptor binding, protein

CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA:

Query Match 78.7%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 3,8e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 3 IEGPTLRQWLARA---GPNIEGPTLRQWLARA 34
 |||||
 Db 1 IEGPTLRQWLARAAGGGGGGIEGPTLRQWLARA 36

RESULT 20

AAB17301

ID AAB17301 standard; Peptide: 36 AA.

AC AAB17301;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:357.

XX Modified peptide: therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

OS WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

PF 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

PI WPI: 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 321; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid

CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA:

Query Match 78.7%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 3,8e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 3 IEGPTLRQWLARA---GPNIEGPTLRQWLARA 34
 |||||
 Db 1 IEGPTLRQWLARAAGGGGGGIEGPTLRQWLARA 36

RESULT 21

AAB17303

ID AAB17303 standard; Peptide: 36 AA.

AC AAB17303;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:359.

XX Modified peptide: therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

OS WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

PF 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

PI WPI: 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 322; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

SQ Sequence 36 AA: 78.7%; Score 144; DB 21; Length 36;
Query Match Best Local Similarity 83.3%; Pred. No. 3.8e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
OY 3 IEGPTLROWLAARA---GPNIGIEGPTLROWLAARA 34
1 IEGPTLROWLAARAAGCGGGIEGPTLROWLAARA 36
Db 1 IEGPTLROWLAARAAGCGGGIEGPTLROWLAARA 36
RESULT 22
AAB17307 standard; Peptide: 36 AA.
ID AAB17307 standard; Peptide: 36 AA.
AC AAB17307;
XX 31-OCT-2000 (first entry)
DT 31-OCT-2000 (first entry)
DE TPO-mimetic peptide sequence SEQ ID NO:363.
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antitumoric; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.
XX Synthetic.
XX WO200024782-A2.
PN 04-MAY-2000.
PD 25-OCT-1999; 99WO-US25044.
PE 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 98US-0428082.
XX (AMGE-) AMGEN INC.
PA Feige U, Liu C, Cheetham J, Boone TC;
PI WPI: 2000-350702/30.
DR Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
PS Example 1; Page 324; 608pp; English.
XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
XX (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
XX independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
XX -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
XX where P1, P2, P3, and P4 = are each independently sequences of
XX pharmacologically active peptides; L1, L2, L3, and L4 = are each
XX independently linkers; and a, b, c, d, e, and f = are each independently
XX 0 or 1, provided that at least 1 of a and b is 1. The composition can
XX have cytostatic, antitumoric, thrombolytic and immunosuppressive
XX activities. DNAs, vectors and host cells from the present invention can
XX be used for producing pharmaceutical compositions. The compositions are
XX useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
XX The use of an Fc domain (rather than a Fab domain) can provide a longer
XX half-life or incorporate functions such as Fc receptor binding, protein
XX A binding, complement fixation, and possibly placental transfer. AAA65943
XX to AAA65926 and AAB16955 to AAB18003 represent nucleotide and amino acid
XX sequences used in the exemplification of the present invention.
XX Sequence 36 AA;

Query Match 78.7%; Score 144; DB 21; Length 36;
Best Local Similarity 83.3%; Pred. No. 3.8e-13;
Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
OY 3 IEGPTLROWLAARA---GPNIGIEGPTLROWLAARA 34
1 IEGPTLROWLAARAAGCGGGIEGPTLROWLAARA 36
Db 1 IEGPTLROWLAARAAGCGGGIEGPTLROWLAARA 36
RESULT 23
AAY96523 standard; Peptide: 36 AA.
ID AAY96523 standard; Peptide: 36 AA.
AC AAY96523;
XX 04-SEP-2000 (first entry)
DT 04-SEP-2000 (first entry)
DE Thrombopoietin mimetic peptide compound 4.
XX Thrombopoietin; mimetic; TMP; platelet; megakaryocyte; production;
XX anti-human immunodeficiency virus; anti-HIV; anti-oncogenic; dermatological;
XX immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX Synthetic.
XX Key Location/Qualifiers
XX Modified-site 1 /note="optionally linked to an Fc molecule"
XX Peptide 1..14 /label= Tmp_1
XX Peptide 15..22 /label= linker
XX Modified-site 18 /note="optionally modified by bromoacetyl or pgs"
XX Peptide 23..36 /label= Tmp_2
PN WO200024770-A2.
PD 04-MAY-2000.
PE 22-OCT-1999; 99WO-US24834.
PR 23-OCT-1998; 98US-0105348.
XX (AMGE-) AMGEN INC.
PA Liu C, Feige U, Cheetham J;
PI WPI: 2000-365108/31.
DR Thrombopoietic peptides which activate mpl receptors and increase the
XX production of platelets or platelet precursors, useful for treatment of
XX diseases which involve thrombocytopenia
PS Claim 16; Page 62; 91pp; English.
XX A compound which binds to an mpl receptor comprising a thrombopoietin
XX mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L1)-nTMP-2]
XX is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
XX 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
XX X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
XX X-1-X-1-4, X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
XX X-4 = P; X-5 = T or S; X-6 = L, I, V, A, F, M, or K; X-11 = A, I, V,
XX or E; X-9 = W, Y or F; X-1-0 = L, I, V, A, F, M, or K; X-1-3 = R, K,
XX L, F, S, T, K, H, or E; X-1-2 = A, I, V, L, F, G, S, or Q; X-1-1 = linker
XX T, V, N, Q or G; X-1-4 = A, I, V, L, F, T, R, E, or G; L-1 = linker
XX comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
XX activate the c-Mpl receptor which mediates the activity of endogenous
XX thrombopoietin. The TMPs are useful for increasing the production of
XX platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
XX is useful for treatment of diseases which involve thrombocytopenia, e.g.
XX aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency

CC virus associated ITP, and systemic lupus erythematosus.
 XX Sequence 36 AA;
 SQ Query Match 78.7%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 3.8e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
 Db 1 IEPTLRQWLARARAGGGGGIEPTLRQWLARA 36
 3 IEPTLRQWLARA---GPNGIEPTLRQWLARA 34
 ||||||| 1 |||||||
 1 IEPTLRQWLARARAGGGGGIEPTLRQWLARA 36
 RESULT 24
 AAY96524
 ID AAY96524 standard; peptide; 36 AA.
 XX AAY96524;
 AC 04-SEP-2000 (first entry)
 DT Thrombopoietin mimetic peptide compound 5.
 DE Thrombopoietin mimetic peptide compound 5.
 XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KM anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
 XX Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT Peptide /note= "optionally linked to an Fc molecule"
 FT Disulfide-bond 1..14
 FT /label= TMP_1
 FT 9..31
 FT Peptide /note= "optional"
 FT 15..22
 FT /label= linker
 FT 23..36
 FT Peptide /label= TMP_2
 FT /label= TMP_2
 PN WO200024770-A2.
 XX 04-MAY-2000.
 PD 22-OCT-1999; 99WO-US24834.
 XX 23-OCT-1998; 98US-0105348.
 PR (AMGE-) AMGEN INC.
 PA
 XX Liu C, Feige U, Cheetham J;
 PI WPI: 2000-365108/31.
 DR Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 PS Claim 16; Page 62; 91pp; English.
 CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X₂-X₁0, X₂-X₁1, X₂-X₁2,
 CC X₂-X₁3, X₂-X₁4, X₁-X₁0, X₁-X₁1, X₁-X₁2, X₁-X₁3, and
 CC X₁-X₁4. X₁=I, A, V, L, S or R; X₂=E, D, K or V; X₃=G or A;
 CC X₄=P; X₅=T or S; X₆=L, I, V, A or F; X₇=R or K; X₈=Q, N,
 CC or E; X₉=W, Y or F; X₁₀=L, I, V, A, F, M, or K; X₁₁=A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂=A, I, V, L, F, G, S, or Q; X₁₃=R, K,
 CC T, V, N, Q or G; X₁₄=A, I, V, L, F, T, R, E, or G; L₁=linker
 CC comprising 10 to 20 amino acids; and n=0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous

CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 SQ Sequence 36 AA;
 Query Match 78.7%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 3.8e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
 Db 1 IEPTLRQWLARARAGGGGGIEPTLRQWLARA 36
 3 IEPTLRQWLARA---GPNGIEPTLRQWLARA 34
 ||||||| 1 |||||||
 1 IEPTLRQWLARARAGGGGGIEPTLRQWLARA 36
 RESULT 25
 AAY96525
 ID AAY96525 standard; peptide; 36 AA.
 XX AAY96525;
 AC 04-SEP-2000 (first entry)
 DT Thrombopoietin mimetic peptide compound 6.
 DE Thrombopoietin mimetic peptide compound 6.
 XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KM anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.
 XX Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT Peptide /note= "optionally linked to an Fc molecule"
 FT 1..14
 FT /label= TMP_1
 FT 15..18
 FT Peptide /label= linker
 FT 19..32
 FT Peptide /label= TMP_2
 FT /label= TMP_2
 FT Modified-site 32
 FT /note= "optionally linked to an Fc molecule"
 PN WO200024770-A2.
 XX 04-MAY-2000.
 PD 22-OCT-1999; 99WO-US24834.
 XX 23-OCT-1998; 98US-0105348.
 PR (AMGE-) AMGEN INC.
 PA
 XX Liu C, Feige U, Cheetham J;
 PI WPI: 2000-365108/31.
 DR Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 XX
 PS Claim 16; Page 62; 91pp; English.
 CC A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)_nTMP_2],
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X₂-X₁0, X₂-X₁1, X₂-X₁2,
 CC X₂-X₁3, X₂-X₁4, X₁-X₁0, X₁-X₁1, X₁-X₁2, X₁-X₁3, and
 CC X₁-X₁4. X₁=I, A, V, L, S or R; X₂=E, D, K or V; X₃=G or A;
 CC X₄=P; X₅=T or S; X₆=L, I, V, A or F; X₇=R or K; X₈=Q, N,
 CC or E; X₉=W, Y or F; X₁₀=L, I, V, A, F, M, or K; X₁₁=A, I, V,

CC L, F, S, T, K, H, or E; X₁-2 = A, I, V, L, F, G, S, or Q; X₁-3 = R, K,
 CC T, V, N, Q or G; X₁-4 = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TmPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX
 SQ Sequence 36 AA:

Query Match 78.7%; Score 144; DB 21; Length 36;
 Best Local Similarity 83.3%; Pred. No. 3.8e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 3 IEGPLRLQWLAAARA---GPNIGIEPTLRQWLAAARA 34
 Db 1 IEGPLRLQWLAAARAAGGGGGGIEPTLRQWLAAARA 36

RESULT 26
 AAB17282
 ID AAB17282 standard; Peptide: 42 AA.

XX
 AC AAB17282;
 XX
 DT 31-OCT-2000 (first entry)
 XX

DE TPO-mimetic peptide sequence SEQ ID NO:338.

XX
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX
 OS Synthetic.
 XX
 PN WO200024782-A2.

XX
 PD 04-MAY-2000.

XX
 PF 25-OCT-1999; 99WO-US25044.

XX
 PR 23-OCT-1998; 98US-0105371.

XX
 PR 22-OCT-1999; 99US-0428082.

XX
 PA (AMGE-) AMGEN INC.

XX
 PI Feige U, Liu C, Cheatham J, Boone TC;
 XX
 DR WPI; 2000-350702/30.

XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Disclosure; Page 313; 608pp; English.

XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X₁)a-F₁-(X₂)b, where: F₁ = an Fc domain; X₁ and X₂ = are each
 CC independently selected from -(L₁)c-p₁-(L₁)c-p₁-(L₂)d-p₂,
 CC -(L₁)c-p₁-(L₂)d-p₂-(L₃)e-p₃, or -(L₁)c-p₁-(L₂)d-p₂-(L₃)e-p₃-(L₄)f-p₄
 CC where p₁, p₂, p₃, and p₄ = are each independently sequences of
 CC pharmacologically active peptides; L₁, L₂, L₃, and L₄ = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can

CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX
 SQ Sequence 42 AA:

Query Match 78.7%; Score 144; DB 21; Length 42;
 Best Local Similarity 83.3%; Pred. No. 4.5e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 3 IEGPLRLQWLAAARA---GPNIGIEPTLRQWLAAARA 34
 Db 1 IEGPLRLQWLAAARAAGGGGGGIEPTLRQWLAAARA 36

RESULT 27
 AAB17311
 ID AAB17311 standard; Peptide: 60 AA.

XX
 AC AAB17311;
 XX
 DT 31-OCT-2000 (first entry)
 XX

DE Synthetic TmP-TmP gene construction peptide SEQ ID NO:385.

XX
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX
 OS Homo sapiens.
 XX
 PN WO200024782-A2.

XX
 PD 04-MAY-2000.

XX
 PF 25-OCT-1999; 99WO-US25044.

XX
 PR 23-OCT-1998; 98US-0105371.

XX
 PR 22-OCT-1999; 99US-0428082.

XX
 PA (AMGE-) AMGEN INC.

XX
 PI Feige U, Liu C, Cheatham J, Boone TC;
 XX
 DR WPI; 2000-350702/30.

XX
 PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 PS Example 2; Page 331; 608pp; English.

XX
 CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X₁)a-F₁-(X₂)b, where: F₁ = an Fc domain; X₁ and X₂ = are each
 CC independently selected from -(L₁)c-p₁-(L₁)c-p₁-(L₂)d-p₂,
 CC -(L₁)c-p₁-(L₂)d-p₂-(L₃)e-p₃, or -(L₁)c-p₁-(L₂)d-p₂-(L₃)e-p₃-(L₄)f-p₄
 CC where p₁, p₂, p₃, and p₄ = are each independently sequences of
 CC pharmacologically active peptides; L₁, L₂, L₃, and L₄ = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AA69443
 CC to AA69526 and AAB1803 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 60 AA;

Query Match 78.7%; Score 144; DB 21; Length 60;
 Best Local Similarity 83.3%; Pred. No. 6.7e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 3 IEPTLRQWLARA---GPNGIEPTLRQWLARA 34
 |||||
 DB 2 IEPTLRQWLARAAGGGGGGIEPTLRQWLARA 37

RESULT 28

AAB16960
 ID AAB16960 standard; Protein; 269 AA.

AC AAB16960;

DT 31-OCT-2000 (first entry)

DE TMP-TMP-Fc protein sequence SEQ ID NO:10.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antitumoral; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

OS Homo sapiens.
 OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PE 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheatham J, Boone TC;

DR WPI: 2000-350702/30.

DR N-PSDB: AAA69446.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 185-186; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antitumoral, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are

CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AA69443
 CC to AA69526 and AAB1803 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 269 AA;

Query Match 78.7%; Score 144; DB 21; Length 269;
 Best Local Similarity 83.3%; Pred. No. 3.6e-12;
 Matches 30; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

OY 3 IEPTLRQWLARA---GPNGIEPTLRQWLARA 34
 |||||
 DB 2 IEPTLRQWLARAAGGGGGGIEPTLRQWLARA 37

RESULT 29

AAB17294
 ID AAB17294 standard; Peptide; 37 AA.

AC AAB17294;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:350.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antitumoral; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PE 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheatham J, Boone TC;

DR WPI: 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 318; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antitumoral, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 37 AA;

Query Match 78.4%; Score 143.5; DB 21; Length 37;
 Best Local Similarity 81.1%; Pred. No. 4.6e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

OY 3 IEPTLRQWLARA-----GPGIEGPTLRQWLARA 34
 ||||||||||||| | |||||||||||||
 Db 1 IEPTLRQWLARAAGGGGGGIEGPTLRQWLARA 37

RESULT 30

AAB17295

ID AAB17295 standard; Peptide; 38 AA.

XX AAB17295;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:351.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; Interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumor necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

OS WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

PF 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

XX WPI: 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

PS Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: P1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)-c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 38 AA;

Query Match 78.1%; Score 143; DB 21; Length 38;
 Best Local Similarity 78.9%; Pred. No. 5.6e-13;
 Matches 30; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

OY 3 IEPTLRQWLARA-----GPGIEGPTLRQWLARA 34
 ||||||||||||| | |||||||||||||
 Db 1 IEPTLRQWLARAAGGGGGGIEGPTLRQWLARA 38

Search completed: October 9, 2002, 08:58:55
 Job time : 15.2881 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using SW model

Run on: October 9, 2002, 08:55:27 ; Search time 5.6534 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-25

Perfect score: 183
Sequence: 1 GGIEGPTLRQWLARARAGPNGIEGPTLRQWLARA 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfilest1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78.5	42.9	25	US-08-764-640-231	Sequence 231, App
2	78.5	42.9	25	US-09-244-298A-231	Sequence 231, App
3	78.5	42.9	25	US-09-516-704-231	Sequence 231, App
4	73	39.9	14	US-08-764-640-13	Sequence 13, App
5	73	39.9	14	US-08-764-640-13	Sequence 13, App
6	73	39.9	14	US-08-764-640-13	Sequence 13, App
7	73	39.9	14	US-08-973-225-13	Sequence 13, App
8	73	39.9	14	US-09-244-298A-13	Sequence 13, App
9	73	39.9	14	US-09-244-298A-13	Sequence 13, App
10	73	39.9	14	US-09-516-704-13	Sequence 13, App
11	73	39.9	14	US-09-516-704-13	Sequence 13, App
12	73	39.9	15	US-08-764-640-17	Sequence 17, App
13	73	39.9	15	US-08-764-640-17	Sequence 17, App
14	73	39.9	15	US-08-973-225-17	Sequence 17, App
15	73	39.9	15	US-08-973-225-17	Sequence 17, App
16	73	39.9	15	US-09-244-298A-17	Sequence 17, App
17	73	39.9	15	US-09-244-298A-17	Sequence 17, App
18	73	39.9	15	US-09-516-704-17	Sequence 17, App
19	73	39.9	15	US-09-516-704-17	Sequence 17, App
20	73	39.9	16	US-08-764-640-18	Sequence 18, App
21	73	39.9	16	US-08-764-640-18	Sequence 18, App
22	73	39.9	16	US-08-764-640-194	Sequence 194, App
23	73	39.9	16	US-08-764-640-232	Sequence 232, App
24	73	39.9	16	US-08-973-225-18	Sequence 18, App
25	73	39.9	16	US-08-973-225-194	Sequence 194, App
26	73	39.9	16	US-09-244-298A-18	Sequence 18, App
27	73	39.9	16	US-09-244-298A-194	Sequence 194, App

28	73	39.9	16	3	US-09-244-298A-232	Sequence 232, App
29	73	39.9	16	4 <td>US-09-516-704-18</td> <td>Sequence 18, App</td>	US-09-516-704-18	Sequence 18, App
30	73	39.9	16	4 <td>US-09-516-704-194</td> <td>Sequence 194, App</td>	US-09-516-704-194	Sequence 194, App
31	73	39.9	16	4 <td>US-09-516-704-232</td> <td>Sequence 232, App</td>	US-09-516-704-232	Sequence 232, App
32	69	37.7	14	2 <td>US-08-764-640-195</td> <td>Sequence 195, App</td>	US-08-764-640-195	Sequence 195, App
33	69	37.7	14	2 <td>US-08-764-640-195</td> <td>Sequence 195, App</td>	US-08-764-640-195	Sequence 195, App
34	69	37.7	14	3 <td>US-08-973-225-195</td> <td>Sequence 195, App</td>	US-08-973-225-195	Sequence 195, App
35	69	37.7	14	3 <td>US-08-973-225-195</td> <td>Sequence 195, App</td>	US-08-973-225-195	Sequence 195, App
36	69	37.7	14	3 <td>US-09-244-298A-195</td> <td>Sequence 195, App</td>	US-09-244-298A-195	Sequence 195, App
37	69	37.7	14	3 <td>US-09-244-298A-195</td> <td>Sequence 195, App</td>	US-09-244-298A-195	Sequence 195, App
38	69	37.7	14	4 <td>US-09-516-704-195</td> <td>Sequence 195, App</td>	US-09-516-704-195	Sequence 195, App
39	69	37.7	14	4 <td>US-09-516-704-199</td> <td>Sequence 199, App</td>	US-09-516-704-199	Sequence 199, App
40	69	37.7	15	2 <td>US-08-764-640-196</td> <td>Sequence 196, App</td>	US-08-764-640-196	Sequence 196, App
41	69	37.7	15	2 <td>US-08-764-640-200</td> <td>Sequence 200, App</td>	US-08-764-640-200	Sequence 200, App
42	69	37.7	15	2 <td>US-08-764-640-209</td> <td>Sequence 209, App</td>	US-08-764-640-209	Sequence 209, App
43	69	37.7	15	2 <td>US-08-764-640-215</td> <td>Sequence 215, App</td>	US-08-764-640-215	Sequence 215, App
44	69	37.7	15	2 <td>US-08-973-225-196</td> <td>Sequence 196, App</td>	US-08-973-225-196	Sequence 196, App
45	69	37.7	15	3 <td>US-08-973-225-200</td> <td>Sequence 200, App</td>	US-08-973-225-200	Sequence 200, App

ALIGNMENTS

RESULT 1
US-08-764-640-231
Sequence 231, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hirshle, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13
Query Match 39.9%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 IEPTLRQWLARA 16
Db 1 IEPTLRQWLARA 14
RESULT 5
US-08-764-640-193
Sequence 193, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193
Query Match 39.9%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 IEPTLRQWLARA 16
Db 1 IEPTLRQWLARA 14
RESULT 6
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherril S.
APPLICANT: Mattheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Magstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A

FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 39.9%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0;

Oy 3 IEPTLRQWLARA 16
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQWLARA 14

RESULT 7
US-08-973-225-193
; Sequence 193, Application US/08973225A
; Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-08-973-225-193

Query Match 39.9%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0;

Oy 3 IEPTLRQWLARA 16
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQWLARA 14

RESULT 8
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palanishappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-Dec-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 39.9%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089; Indels 0; Gaps 0;
Matches 14; Conservative 0; Mismatches 0;

Oy 3 IEPTLRQWLARA 16
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQWLARA 14

RESULT 9
US-09-244-298A-193
; Sequence 193, Application US/09244298A

Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-193
Query Match 39.9%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 IEQPTLROWLAARA 16
DB 1 IEQPTLROWLAARA 14
RESULT 10
US-09-516-704-13
Sequence 13, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13
Query Match 39.9%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 IEQPTLROWLAARA 16
DB 1 IEQPTLROWLAARA 14
RESULT 11
US-09-516-704-193
Sequence 193, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match 39.9%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00089;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IEGPLRLQWLAARA 16
|||||
DB 1 IEGPLRLQWLAARA 14

RESULT 12

US-08-764-640-17
Sequence 17, Application US/08764640
Patent No. 5869451 5837683

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-17

Query Match 39.9%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IEGPLRLQWLAARA 16
|||||
DB 1 IEGPLRLQWLAARA 14

RESULT 13

US-08-764-640-185
Sequence 185, Application US/08764640
Patent No. 5869451 5837683

GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-185

Query Match 39.9%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IEGPLRLQWLAARA 16
|||||
DB 2 IEGPLRLQWLAARA 15

RESULT 14
US-08-973-225-17
Sequence 17, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwila, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherill S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK30650SW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-973-225-17
Query Match 39.9%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. NO. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 IEGPLRLRWLAARA 16
Db 1 IEGPLRLRWLAARA 14
RESULT 15
US-08-973-225-185
Sequence 185, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwila, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherill S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK30650SW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185
Query Match 39.9%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. NO. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 IEGPLRLRWLAARA 16
Db 2 IEGPLRLRWLAARA 15

Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK30650SW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185
Query Match 39.9%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. NO. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 IEGPLRLRWLAARA 16
Db 2 IEGPLRLRWLAARA 15
RESULT 16
US-09-244-298A-17
Sequence 17, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwila, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniasappan
Wagstrom, Christopher R.
Hendren, Richard W.
Depinice, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK30650SW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185
Query Match 39.9%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. NO. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 IEGPLRLRWLAARA 16
Db 2 IEGPLRLRWLAARA 15

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17

Query Match 39.9%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLQWLAARA 16
Db 1 IEPTLQWLAARA 14

RESULT 17
US-09-244-298A-185
Sequence 185, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagsstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduri, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-185

Query Match 39.9%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLQWLAARA 16
Db 2 IEPTLQWLAARA 15

RESULT 18
US-09-516-704-17
Sequence 17, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirila, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagsstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduri, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match 39.9%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00096;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLQWLAARA 16
Db 1 IEPTLQWLAARA 14

Db 1 IEGLTROWLAARA 14

RESULT 19

US-09-516-704-185

Sequence 185, Application US/09516704

Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Poddaturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubic, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-09-516-704-185

Query Match 39.9%; Score 73; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.00096;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IEGLTROWLAARA 16

Db 2 IEGLTROWLAARA 15

RESULT 20

US-08-764-640-18

Sequence 18, Application US/08764640

Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Poddaturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

APPLICANT: Balasubramanian, Palaniappan

APPLICANT: Wagstrom, Christopher R.

APPLICANT: Hendren, Richard W.

APPLICANT: Deprince, Randolph B.

APPLICANT: Poddaturi, Surekha

APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubic, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product="Beta-ala"

US-08-764-640-18

Query Match 39.9%; Score 73; DB 2; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.001;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IEGLTROWLAARA 16

Db 1 IEGLTROWLAARA 14

RESULT 21

US-08-764-640-194

Sequence 194, Application US/08764640

Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Poddaturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-194

Query Match 39.9%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQWLARA 15

RESULT 22
US-08-764-640-232
Sequence 232, Application US/08764640
Patent No. 5869451
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depriuce, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-232

Query Match 39.9%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQWLARA 15

RESULT 23
US-08-973-225-18
Sequence 18, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherril S.
APPLICANT: Matheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK306505M
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product="Beta-ala"
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18

Query Match 39.9%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
|||||
Db 1 IEPTLRQWLARA 14

RESULT 24

US-08-973-225-194
Sequence 194, Application US/08973225A
Patent No. 6083913

GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Magstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESSES:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>

MOLECULE TYPE: peptide
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194

Query Match 39.9%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
|||||
Db 2 IEPTLRQWLARA 15

RESULT 25

US-08-973-225-220
Sequence 220, Application US/08973225A
Patent No. 6083913

GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Magstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESSES:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 220:

SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>

MOLECULE TYPE: peptide
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220

Query Match 39.9%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
|||||
Db 2 IEPTLRQWLARA 15

RESULT 26

US-09-244-298A-18
Sequence 18, Application US/09244298A
Patent No. 6121238

GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Magstrom, Christopher R.
Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR

NUMBER OF SEQUENCES: 232

CORRESPONDENCE ADDRESSES:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 220:

SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>

MOLECULE TYPE: peptide
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220

Query Match 39.9%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
|||||
Db 2 IEPTLRQWLARA 15

APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "beta-ala"
US-09-244-298A-18

Query Match 39.9%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
| | | | | | | | | | | | | | | | | |
DB 1 IEPTLRQWLARA 14

RESULT 27
US-09-244-298A-194
Sequence 194, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depina, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 39.9%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
| | | | | | | | | | | | | | | | | |
DB 2 IEPTLRQWLARA 15

RESULT 28
US-09-244-298A-232
Sequence 232, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depina, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-232

Query Match 39.9%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
DB 2 IEPTLRQWLARA 15

RESULT 29
US-09-516-704-18
Sequence 18, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Depinice, Randolph B.
Podatuturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE: NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Beta-ala"
SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-516-704-18

Query Match 39.9%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
DB 1 IEPTLRQWLARA 14

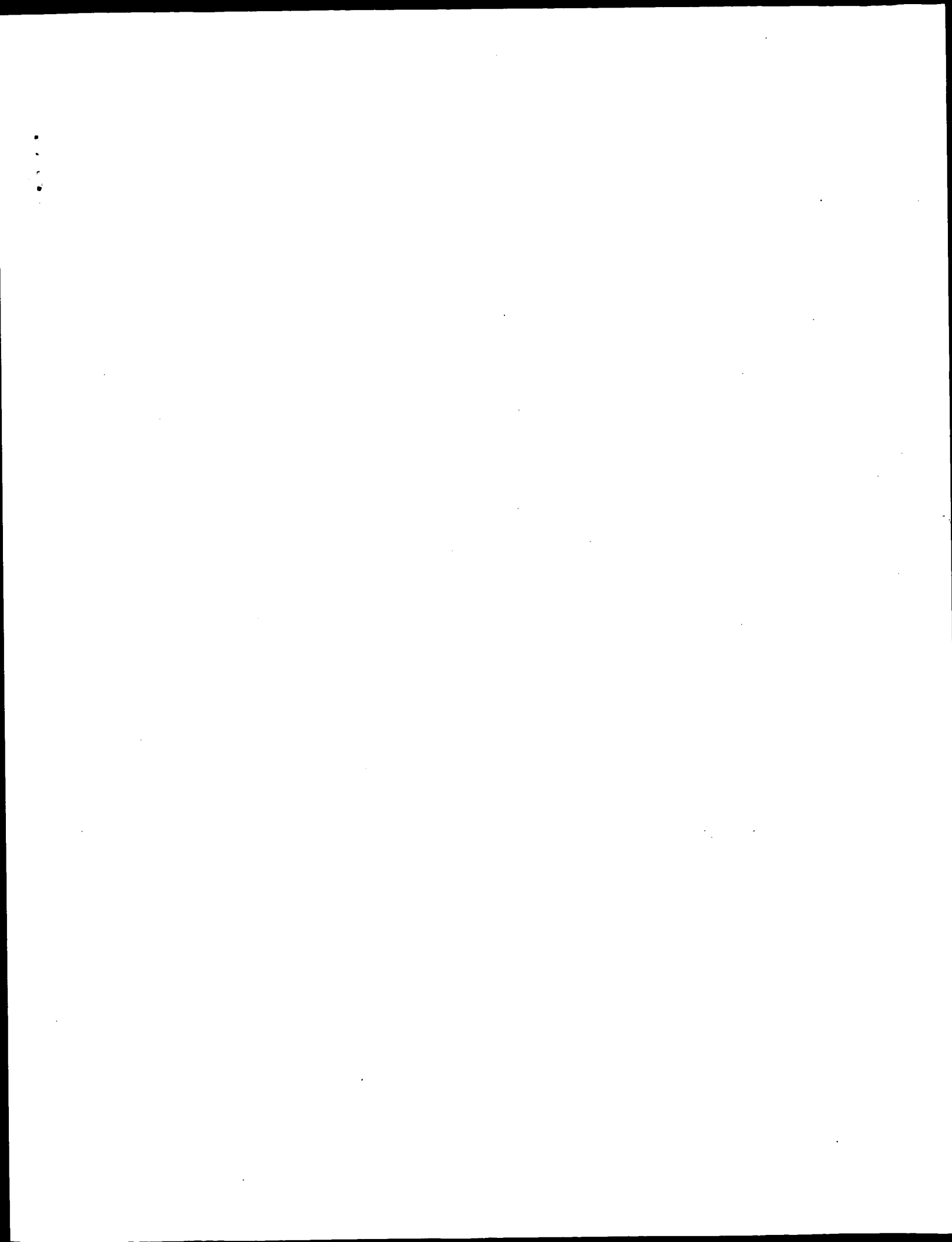
RESULT 30
US-09-516-704-194
Sequence 194, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Depinice, Randolph B.
Podatuturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 194:

US-09-516-704-194

Query Match 39.9%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IEPTLRQWLARA 16
DB 2 IEPTLRQWLARA 15

Search completed: October 9, 2002, 09:06:30
Job time : 5.6534 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 : Search time 7.64403 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838C-25

Perfect score: 183

Sequence: 1 GGIEGPTLRQWLARAGPNEGIEPTLRQWLAR 34

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	34.4	683	2 B71325	conserved hypotet
2	58	31.7	214	2 T22896	hypothetical prote
3	56	30.6	361	2 F87286	cation efflux fami
4	55	30.1	403	2 AD0748	lysine-specific
5	55	30.1	420	2 JMO076	sorbitol oxidase (
6	55	30.1	440	2 S65358	familial Alzheimer
7	54	29.5	346	2 D85818	unknown protein en
8	54	29.5	1095	2 B83471	probable pyruvate
9	54	29.5	1366	1 CGH025	collagen alpha 2(I
10	54	29.5	3198	2 A43426	collagen alpha 2(I
11	53.5	29.2	371	2 F83487	hypothetical prote
12	53	29.0	296	2 AG0147	probable membrane
13	53	29.0	326	2 C75350	probable UV damage
14	53	29.0	524	1 VGVNVC	spike glycoprotein
15	53	29.0	1373	1 A43291	collagen alpha 2(I
16	53	29.0	1433	2 A46053	bulbosus pemphigoid
17	52.5	28.7	814	2 G02390	disintegrin-like m
18	52	28.4	214	2 T22892	hypothetical prote
19	52	28.4	230	2 T32999	hypothetical prote
20	52	28.4	356	2 T35254	conserved hypotet
21	51.5	28.1	150	2 AF3634	nitric-oxide reduc
22	51.5	28.1	621	2 AH2257	hypothetical prote
23	51	27.9	215	2 T22895	hypothetical prote
24	51	27.9	246	2 AH0190	probable oxidoredu
25	51	27.9	281	2 G72680	hypothetical prote
26	51	27.9	306	2 D70601	UTP--glucose-1-pho
27	51	27.9	589	2 T29299	hypothetical prote
28	51	27.9	600	2 C83221	transport protein
29	51	27.9	697	1 S04987	SITS-binding prote

30	51	27.9	719	2 B95325	conserved hypotet
31	51	27.9	1838	1 CGH01V	collagen alpha 1(V
32	51	27.9	1843	2 S18803	collagen alpha 1(V
33	50.5	27.6	904	2 C70559	probable pola prot
34	50	27.3	207	2 B75327	hypothetical prote
35	50	27.3	298	2 T32371	hypothetical prote
36	50	27.3	351	2 C75479	conserved hypotet
37	50	27.3	410	1 DEPSXA	3-methyl-2-oxobuta
38	50	27.3	38	2 C83365	2-oxoisovalerate d
39	50	27.3	415	2 T38324	probable trna meth
40	50	27.3	460	2 S06469	photosystem II chl
41	50	27.3	472	2 T20454	hypothetical prote
42	50	27.3	1446	1 A45344	immediate-early pr
43	49.5	27.0	333	2 A36925	transcription acti
44	49.5	27.0	341	2 A13083	monooxygenase [imp
45	49.5	27.0	355	2 H98202	hypothetical prote

ALIGNMENTS

RESULT 1

B71325 conserved hypotetrical protein TP0421 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: B71325
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: B71325
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-683 <COL>
A:Cross-references: GB:AE001220; GB:AE000520; NID:93322705; PIDN:AAC5409.1; PID:93332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0421

Query Match 34.4% Score 63; DB 2; Length 683;
Best Local Similarity 46.4% Pred. No. 3.2;
Matches 13; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

OY 6 PTLRQWLARAGPNEGIEPTLRQWLAR 33

Db 74 PTLRQWLARAGPNEGIEPTLRQWLAR 101

RESULT 2

T22896 hypothetical protein F58B3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T22896

R:Harris, B. submitted to the EMBL Data Library, May 1996

A:Reference number: Z19633

A:Accession: T22896

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-214 <WIL>

A:Cross-references: EMBL:Z73427; PIDN:CAA97801.1; GSPDB:GN00022; CESP:F58B3.3

A:Experimental source: clone F58B3

C:Genetics:
A:Gene: CESP:F58B3.3

A:Map position: 4

A:Introns: 68/1

C:Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match 31.7% Score 58; DB 2; Length 214;

Best Local Similarity 47.6%; Pred. No. 4;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
QY 1 GGIEGPTLRQWLARAGPNCI 21
DB 186 GGMSPTIRHOMESTTGAPCGV 206

RESULT 3

F87286
cation efflux family protein [imported] - *Caulobacter crescentus*
C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: F87286
R:Metman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <STO>

A:Cross-references: GB:AE005673; NID:g13421446; PIDN:AAK22290.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0303

Query Match 30.6%; Score 56; DB 2; Length 361;
Best Local Similarity 54.5%; Pred. No. 12;
Matches 12; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 12 LARAGPNCIEGPTLRQWLAR 33
DB 266 LALDAPRCIDYQKVRDWLAR 287

RESULT 4

AD0748
tyrosine-specific transport protein STY2145 [imported] - *Salmonella enterica* subsp. *enterica*
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh
A:Note: This species has also been called *Salmonella typhi*

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD0748
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
Th. T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AD0748
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-403 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05687.1; PID:g16503181; GSPDB:GN00176
C:Genetics:
A:Gene: STY2145
C:Superfamily: tyrosine-specific transport protein

Query Match 30.1%; Score 55; DB 2; Length 403;
Best Local Similarity 52.4%; Pred. No. 18;
Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 GGIEGPTLRQWLARAGPNCI 21
DB 238 GSIDSPTRGLASHAGNGL 258

RESULT 5
JM0076
sorbitol oxidase (EC 1.1.1.1) - *Streptomyces* sp.
C:Species: *Streptomyces* sp.

C:Date: 17-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 24-Oct-2000
C:Accession: JM0076
R:Hiraga, K.; Eto, T.; Yoshioke, T.; Oda, K.
Biosci. Biotechnol. Biochem. 62, 347-353, 1998
A:Title: Molecular cloning and expression of a gene encoding a novel sorbitol oxidase
A:Reference number: JM0076; MUID:98193986
A:Accession: JM0076
A:Molecule type: mRNA
A:Residues: 1-420 <HTR>

A:Cross-references: DDBJ:AB000519; NID:g1856966; PIDN:BA19135.1; PID:g1856967
C:Comment: This protein oxidizes D-sorbitol to produce hydrogen peroxide and glucose
C:Superfamily: L-gulonolactone oxidase
C:Keywords: oxidoreductase

Query Match 30.1%; Score 55; DB 2; Length 420;
Best Local Similarity 37.9%; Pred. No. 19;
Matches 11; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 5 GPTLRQWLARAGPNCIEGPTLRQWLAR 33
DB 215 GPVGQVWLKQKRVDEGARSVPAEWLGAR 243

RESULT 6

S65358
familial Alzheimer's disease protein 1 - human
C:Species: *Homo sapiens* (man)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 25-Apr-1997

C:Accession: S65358
R:Matsumoto, A.; Matsumoto, R.; Fujiwara, Y.
Eur. J. Biochem. 230, 337-343, 1995
A:Title: Molecular cloning of human CDNA with a sequence highly similar to that of th
A:Reference number: S65358; MUID:95324544
A:Accession: S65358
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-440 <MAT>

Query Match 30.1%; Score 55; DB 2; Length 440;
Best Local Similarity 45.5%; Pred. No. 20;
Matches 15; Conservative 1; Mismatches 11; Indels 6; Gaps 1;

QY 2 GIEGPTLRQWLARAGPNCIEGPTLRQWLARA 34
DB 371 GERGPDLRSALAGRVPTGP-----PFSARA 397

RESULT 7

D85818
unknown protein encoded within prophage CP-933U [imported] - *Escherichia coli* (strain
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: D85818
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
hiller, L.; Grodbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: D85818
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>

A:Cross-references: GB:AE005174; NID:g12516109; PIDN:AG57008.1; GSPDB:GN00145; UMGCP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3092

Query Match 29.5%; Score 54; DB 2; Length 346;
Best Local Similarity 41.9%; Pred. No. 20;
Matches 13; Conservative 3; Mismatches 9; Indels 6; Gaps 1;

QY 5 GPTLRQWLARAG-----PNCIEGPTLRQW 29

DB 2 GDCIROMLARAAGFENVERKDNANGMTLREW 32

RESULT 8
B83471
probable pyruvate carboxylase PA1400 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83471
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, N.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: B83471
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1095 <STO>
A:Cross-references: GB:AE004569; GB:AE004091; NID:g9947339; PIDN:AAG04789.1; GSPDB:GN001
A:Experimental source: strain PA01
A:Genetics: PA1400

Query Match 29.5%; Score 54; DB 2; Length 1095;
Best Local Similarity 45.5%; Pred. No. 68;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 3 IECPTRQMLAARAGPNGIECP 24
||| | : | | | : |||
Db 786 IEQGLGRFAAEVGPPTGVGP 807

RESULT 9
CGH025
collagen alpha 2(I) chain precursor - human
N:Alternate names: procollagen alpha 2(I) chain
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1989 #sequence_revision 25-Aug-1995 #text_change 21-Jul-2000
C:Accession: A28500; S00824; S09176; I55311; A58111; A28472; A42165; A34405; A90567; I553005; A02865
R:de Wet, W.; Bernard, M.; Benson-Chanda, V.; Chu, M.L.; Dickson, L.; Weil, D.; Ramirez, J. Biol. Chem. 262, 16032-16036, 1987
A:Title: Organization of the human pro-alpha-2(I) collagen gene.
A:Reference number: A28500; MUID:88058962
A:Accession: A28500
A:Molecule type: DNA; mRNA
A:Residues: 1-248, 'N', 250-1366 <DEW>
A:Cross-references: GB:J03464; NID:g179595; PIDN:AAB59374.1; PID:g179596
R:Kuivaniemi, H.; Tromp, G.; Chu, M.L.; Prockop, D.J.
Biochem. J. 252, 633-640, 1988
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-2(I) chain of human alpha-2(I) procollagen
A:Reference number: S00824; MUID:88339824
A:Accession: S00824
A:Molecule type: mRNA
A:Residues: 1-275, 'A', 277-332, 'V', 334-337, 'A', 339-482, 'A', 484-548, 'D', 550-765 <KU11>
A:Cross-references: EMBL:X00724; NID:g30022; PIDN:CA68709.1; PID:g30023
R:Dickson, L.A.; de Wet, W.; di Liberto, M.; Weil, D.; Ramirez, F.
Nucleic Acids Res. 13, 3427-3438, 1985
A:Title: Analysis of the promoter region and the N-propeptide domain of the human proalpha-2(I) procollagen gene
A:Reference number: S09176; MUID:85242047
A:Accession: S09176
A:Molecule type: DNA
A:Residues: 1-23, 33-58, 'P', 60-93 <DIC>
A:Cross-references: EMBL:X02488; NID:g30098; PIDN:CAA26320.1; PID:g30099
R:Weil, D.; D'Alessio, M.; Ramirez, F.; Eyre, D.R.
J. Biol. Chem. 265, 16007-16011, 1990
A:Title: Structural and functional characterization of a splicing mutation in the pro-alpha-2(I) procollagen gene
A:Reference number: I55311; MUID:90368825
A:Accession: I55311
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 76-93 <WE11>

A:Cross-references: GB:M35391; NID:g189684; PIDN:AAA60041.1; PID:g189685
A:Accession: A58111
A:Molecule type: protein
A:Residues: 23-75, 94-96 <WE12>
A>Note: mutant sequence from a patient with Ehlers-Danlos syndrome type VII
R:Wirtz, M.K.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, D.W.
J. Biol. Chem. 262, 16376-16385, 1987
A:Title: Ehlers-Danlos syndrome type VII. Deletion of 18 amino acids comprising the propeptide
A:Reference number: A28472; MUID:88059013
A:Accession: A28472
A:Molecule type: protein
A:Residues: 32-75, 94-111 <WIR>
A>Note: mutant sequence of patient with Ehlers-Danlos syndrome type VII
R:Chiodo, A.A.; Hockley, A.; Cole, W.G.
J. Biol. Chem. 267, 6361-6369, 1992
A:Title: A base substitution at the splice acceptor site of intron 5 of the COL1A2 gene
S-Danlos syndrome type VII.
A:Reference number: A42165; MUID:92210617
A:Accession: A42165
A:Molecule type: mRNA
A:Residues: 50-126 <CHI>
A>Note: parts of this sequence were determined by protein sequencing; a mutant sequence
R:Weil, D.; D'Alessio, M.; Ramirez, F.; Steinmann, B.; Wirtz, M.K.; Glanville, R.W.;
J. Biol. Chem. 264, 16804-16809, 1989
A:Title: Temperature-dependent expression of a collagen splicing defect in the fibroblast
A:Reference number: A34405; MUID:89380311
A:Accession: A34405
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 58-108 <WE13>
A:Cross-references: GB:J05049
A>Note: the accession cited by the authors is not found in Genbank
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha-2(I) procollagen gene
A:Reference number: A90567; MUID:71038625
A:Accession: A90567
A:Molecule type: protein
A:Residues: 7, 81, 'B', 83-96; 417-447 <CLR>
A>Note: the compositions of peptides CNBR-1, CNBR-0, and CNBR-2 were determined; evidence
R:Kuivaniemi, H.; Sabol, C.; Tromp, G.; Sipola-Thiele, M.; Prockop, D.J.
J. Biol. Chem. 263, 11407-11413, 1988
A:Title: A 19-base pair deletion in the pro-alpha 2(I) gene of type I procollagen that is asymptomatic mother.
A:Reference number: I55264; MUID:88298792
A:Accession: I55264
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA; mRNA
A:Residues: 145-197 <KU12>
A:Cross-references: GB:M21671; NID:g189521; PIDN:AAA59994.1; PID:g553606
A>Note: single base mutation in intron leads to abnormal splicing of mRNA
R:Chapman, S.D.; Shapiro, J.R.; McKinstry, M.B.; Stover, M.L.; Branson, P.; Rowe, D.W.
J. Bone Miner. Res. 7, 793-805, 1992
A:Title: Expression of mutant alpha 1(I)-procollagen in osteoblast and fibroblast cultures
A:Reference number: I55485; MUID:92351816
A:Accession: I55485
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 163-181, 200-213 <CH2>
A:Cross-references: GB:S41099; NID:g252702; PIDN:AAB22761.1; PID:g252703
A>Note: mutant sequence from a patient with osteogenesis imperfecta type IV
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagen
A:Reference number: A92069; MUID:71001508
A:Accession: A92069
A:Molecule type: protein
A:Residues: 175-180 <WOR>
A:Experimental source: skin
R:Fietzek, P.P.; Furtmayer, H.; Kuehn, K.
Eur. J. Biochem. 47, 257-261, 1974

A:Title: Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig-ski
 A:Reference number: A91224; MUID:75008198
 A:Accession: A91224
 A:Molecule type: protein
 A:Residues: 418-447 <FLD>
 R:Triomp, G.; Prockop, D.J.
 Proc. Natl. Acad. Sci. U.S.A. 85, 5254-5258, 1988
 A:Title: Single base mutation in the pro alpha 2(I) collagen gene that causes efficient
 A:Reference number: 159125; MUID:88276936
 A:Accession: 159125
 A>Status: translation not shown; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 520-573 <TRC>
 A:Cross-references: GB:M2153; NID:q180881; PIDN:AAA52053.1; PID:q180882
 A:Note: single base mutation in intron leads to splicing out of exon 28
 R:Bernard, M.P.; Myers, J.C.; Chu, M.L.; Ramirez, F.; Elkenberry, E.F.; Prockop, D.J.
 Biochemistry 22, 1139-1145, 1983
 A:Title: Structure of a cDNA for the proalpha-2 chain of human type I procollagen. Compa
 A:Reference number: S09174; MUID:83178919
 A:Accession: S09174
 A:Molecule type: mRNA
 A:Residues: 623-742, 'A', 744-764, 'X', 766-827, 'A', 829-830, 'P', 832-836, 'P', 838-1097, 'L', 109
 A:Cross-references: GB:J00115; GB:V00503; NID:930123; PIDN:CAA23761.1; PID:9825646
 A:Experimental source: skin fibroblast cells
 R:Forlino, A.; Zolazzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Motte
 Hum. Mol. Genet. 3, 2201-2206, 1994
 A:Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
 A:Reference number: 154365; MUID:95187161
 A:Accession: 154365
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 663-675, 'V', 677, 'P', 679-742, 'A', 744-746 <FOR>
 A:Cross-references: GB:LA7668; NID:q1009095; PIDN:AA559577.1; PID:q1009096
 R:Niyidizi, C.; Bondio, J.; Byers, P.H.; Eyre, D.R.
 J. Biol. Chem. 267, 23108-23112, 1992
 A:Title: Incorporation of type I collagen molecules that contain a mutant alpha 2(I) cha
 A:Reference number: 155369; MUID:93054657
 A:Accession: 155369
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 665-666, 'D', 668-670 <NIT>
 A:Cross-references: GB:U00613; NID:q180888; PIDN:AA55934.1; PID:q180889
 A:Note: mutant sequence from a patient with osteogenesis imperfecta
 R:Bateman, J.F.; Hannagan, M.; Chan, D.; Cole, W.G.
 Biochem. J. 276, 765-770, 1991
 A:Title: Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitut
 e method.
 A:Reference number: A56799; MUID:91291136
 A:Accession: A56799
 A:Molecule type: mRNA
 A:Residues: 672-675, 'V', 677, 'P', 679-681 <BAT>
 A:Cross-references: GB:S39878; NID:q1679911; PIDN:AA19114.1; PID:q232761
 A:Experimental source: fibroblast cell culture
 A:Note: sequence extracted from NCBI backbone (NCBIN:39878, NCBI:39886)
 A:Note: mutant sequence of patient with osteogenesis imperfecta type IV; the authors sug
 nstrol sequence
 R:Maekelae, J.K.; Vuorio, T.; Vuorio, E.
 Biochim. Biophys. Acta 1049, 171-176, 1990
 A:Title: Growth-dependent modulation of type I collagen production and mRNA levels in cu
 A:Reference number: S10768; MUID:90304220
 A:Accession: S10768
 A:Molecule type: mRNA
 A:Residues: 960-1021, 'L', 1023-1188, 'D', 1190-1197, 'S', 1199-1356 <MAE>
 A:Cross-references: EMBL:X55525; NID:930101; PIDN:CAA39142.1; PID:930102
 A:Experimental source: fibroblast cell culture
 R:Wyers, J.C.; Chu, M.L.; Faro, S.H.; Clark, W.J.; Prockop, D.J.; Ramirez, F.
 Proc. Natl. Acad. Sci. U.S.A. 78, 3516-3520, 1981
 A:Title: Cloning a cDNA for the pro-alpha2 chain of human type I collagen.
 A:Reference number: A18855; MUID:81273090
 A:Accession: A18855
 A:Molecule type: mRNA
 A:Residues: 964-979, 'V', 981-1018, 'Q', 1020 <MYE>
 A:Cross-references: GB:J00114; NID:9180393; PIDN:AA51996.1; PID:q180394

A:Note: 1019-Lau was also found
 R:Wenstrup, R.J.; Cohn, D.H.; Cohen, T.; Byers, P.H.
 J. Biol. Chem. 263, 7734-7740, 1988
 A:Title: Arginine for glycine substitution in the triple-helical domain of the produc
 A:Reference number: 155285; MUID:88227975
 A:Accession: 155285
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1090-1107 <MEN1>
 A:Cross-references: GB:M22816; NID:q179602; PIDN:AAA51844.1; PID:q179603
 A:Accession: 170059
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1090-1101, 'R', 1103-1107 <MEN2>
 A:Cross-references: GB:M22817; NID:q179606; PIDN:AAA51846.1; PID:q179607
 A:Note: mutant sequence from a patient with osteogenesis imperfecta type IV
 R:Myers, J.C.; Dickson, L.A.; de Wet, W.J.; Bernard, M.P.; Chu, M.L.; di Liberto, M.;
 J. Biol. Chem. 258, 10128-10135, 1983
 A:Title: Analysis of the 3' end of the human pro-alpha-2(I) collagen gene. Utilizatio
 A:Reference number: S09175; MUID:83290853
 A:Accession: S09175
 A:Molecule type: DNA

Query Match 29.5%; Score 54; DB 1; Length 1366;
 Best Local Similarity 52.2%; Pred. No. 85;
 Matches 12; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 GIEPTLRQWLAARAGPNCIECP 24
 :||| | ||||| ||
 Db 751 GVGPTGPVGAAGPAGPAGP 773

RESULT 10

collagen alpha 2 fibrillar chain precursor - sea urchin (Strongylocentrotus purpuratus
 C:Species: Strongylocentrotus purpuratus (purple urchin)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
 C:Accession: A43426
 R:Exposito, J.Y.; D'Alessio, M.; Ramirez, F.
 J. Biol. Chem. 267, 17404-17408, 1992
 A:Title: Novel amino-terminal propeptide configuration in a fibrillar procollagen und
 A:Reference number: A43426; MUID:92381062
 A:Accession: A43426
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-3198 <EXP>
 A:Cross-references: GB:M92041; NID:q161448; PIDN:AAA30040.1; PID:q161449
 A:Note: sequence extracted from NCBI backbone (NCBI:111965)
 C:Superfamily: unassigned collagens; fibrillar collagen carboxyl-terminal homology; v
 F:48-106/Domain: von Willebrand factor type C repeat homology <WVC>
 F:2978-3198/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 29.5%; Score 54; DB 2; Length 3198;
 Best Local Similarity 44.8%; Pred. No. 2; Le+02;
 Matches 13; Conservative 2; Mismatches 8; Indels 6; Gaps 1;

QY 2 GIEPTLRQWLA-----ARAGPNCIECP 24
 :||| | ||||| ||
 Db 1843 GVGPTGPVGAAGPAGPAGP 1871

RESULT 11

F83487
 hypothetical protein PA1267 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83487
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mitsuoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; L
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337

A:Accession: F83487
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-371 <STO>
A:Cross-references: GB:AE004556; GB:AE004091; NID:g947194; PIDN:AA604656.1; GSPDB:GN001
A:Experimental source: strain PRO1
C:Genetics:
A:Gene: PA1267

Query Match 29.2%; Score 53.5; DB 2; Length 371;
Best Local Similarity 29.8%; Pred. No. 25;
Matches 14; Conservative 8; Mismatches 12; Indels 13; Gaps 2;

OY 1 GIEGPTLRQMLAARAGP-----NGIEGPTLR-----QMLARA 34
DB 140 GILYAPNARWLLDAGPRLRLVAYEVDGSRRLADGRWLAE 186

RESULT 12

AG0147
Probable membrane protein YPO1203 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AG0147
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tilhail, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; 11; M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:1156360
A:Accession: AG0147
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-296 <KUR>
A:Cross-references: GB:AUS90842; PIDN:CAC90042.1; PID:g15979263; GSPDB:GN00175
C:Genetics:
A:Gene: YPO1203

Query Match 29.0%; Score 53; DB 2; Length 296;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 16; Conservative 1; Mismatches 11; Indels 4; Gaps 1;

OY 3 IEPTLRQMLAARAGPNGIEGPTLRQMLARA 34
DB 49 IAGVLFPLAIR-----GHALPTLRQMAAASA 76

RESULT 13

C75350
Probable UV damage endonuclease - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: C75350
R:White, O.; Eichen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <WHI>
A:Cross-references: GB:AE002022; GB:AE000513; NID:g6459590; PIDN:AAFI1370.1; PID:g645959
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1819
A:Map position: 1

Query Match 29.0%; Score 53; DB 2; Length 326;
Best Local Similarity 40.5%; Pred. No. 26;
Matches 15; Conservative 4; Mismatches 6; Indels 12; Gaps 2;

OY 4 EGPTLRQMLAARAGP-----PNCIEGPTLRQ 28
DB 249 EDSVREHVIARATWQPREMOWHLNNGIEGPDOR 285

RESULT 14

GVNVCV
spike glycoprotein G precursor - Chandipura virus

C:Species: Chandipura virus
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999
C:Accession: A32443
R:McIntosh, P.S.; Bello, R.S.; Butcher, M.; Patel, B.; Ghosh, H.P.; Banerjee, A.K. Virology 171, 285-290, 1989

A:Title: Structure and expression of the glycoprotein gene of Chandipura virus.
A:Reference number: A32443; MUID:89299473
A:Accession: A32443
A:Molecule type: mRNA
A:Residues: 1-524 <MAS>
A:Cross-references: GB:J04350; NID:g323376; PIDN:AAA42916.1; PID:g323377
C:Genetics:
A:Gene: G
C:Superfamily: rhabdovirus spike glycoprotein G
C:Keywords: glycoprotein; spike protein; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-524/Product: spike glycoprotein G #status predicted <SGG>
F:472-491/Domain: transmembrane #status predicted <TM>
F:164,344/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.0%; Score 53; DB 1; Length 524;
Best Local Similarity 37.0%; Pred. No. 42;
Matches 10; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

OY 3 IEPTLRQMLAARAGPNGIEGPTLRQ 29
DB 359 IDGPVLEPKKRESPPSSISDITQW 385

RESULT 15

A43291
collagen alpha 2(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A43291; A54328
R:Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J. Genomics 13, 1345-1346, 1992

A:Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collag
A:Reference number: A43291; MUID:92372043
A:Accession: A43291
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1373 <PHI>
A:Cross-references: GB:X58251; NID:g50488; PIDN:CAA41205.1; PID:g50489
A:Note: sequence extracted from NCBI backbone (NCBI:P:112027)
R:Phillips, C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J. J. Invest. Dermatol. 97, 980-984, 1991
A:Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polym
A:Reference number: A54328; MUID:92084969
A:Accession: A54328
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-110 <PHI>
C:Genetics:
A:Gene: COL1A2
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:1145-1373/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 29.0%; Score 53; DB 1; Length 1373;
Best Local Similarity 52.2%; Pred. No. 11e+02;
Matches 12; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

OY 2 GIEGPTLRQMLAARAGPNGIEG 24
II IIII I :IIII II

Db 757 GIVGPTGSGVAGAPSGPNCPPGP 779

RESULT 16

A46053
 bullous pemphigoid antigen, BPAG2, type XVII collagen alpha 1-chain - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 21-Sep-1993 #sequence_revision 21-Sep-1993 #text_change 05-Nov-1999
 C:Accession: A46053
 R:Li, K.; Tamai, K.; Tan, E.M.L.; Uitto, J.
 J. Biol. Chem. 268, 8825-8834, 1993
 A:Title: Cloning of type XVII collagen. Complementary and genomic DNA sequences of mouse segment, and unusual features in the 5'-end of the gene and the 3'-untranslated region of A:Reference number: A46053; MUID:93232041
 A:Accession: A46053
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1433
 A:Cross-references: GB:I08407; NID:9309182; PID:AAA37443.1; PID:9309183
 A:Note: sequence extracted from NCBI backbone (NCBIN:129627, NCBI:P:129628)

Query Match

Best Local Similarity 29.0%; Score 53; DB 2; Length 1433;
 Pred. No. 1.2e+02;
 Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 2 GIEGPTLRQWLARAGPNCIEGP 24

Db 622 GMEGPTRGRLAGPMPRGP 644

RESULT 17

G02390
 disintegrin-like metalloproteinase MDC15 (EC 3.4.24.-) - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 31-Dec-2000
 C:Accession: G02390; PC4263
 R:Herren, B.; Raines, E.W.; Ross, R.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: H01157
 A:Accession: G02390
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-814 <HE>
 A:Cross-references: EMBL:U46005; NID:91335871; PID:MAC51112.1; PID:91335872
 R:McKie, N.; Edwards, T.; Dallas, D.J.; Houghton, A.; Stringer, B.; Graham, R.; Russell, B.
 Biochem. Biophys. Res. Commun. 230, 335-339, 1997
 A:Title: Expression of members of a novel membrane linked metalloproteinase family (ADAM A:Reference number: PC4263; MUID:97168971
 A:Accession: PC4263
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-461 <MCK>
 A:Experimental source: articular chondrocyte
 C:Comment: This protein is a membrane bound protein and involved in cell/cell and cell/m C:Superfamily: mouse meltrin alpha; disintegrin homology
 C:Keywords: hydrolase; metalloproteinase; zinc
 F:420-503/Domain: disintegrin homology <DIS>
 F:348-352/Binding site: zinc (His) #status predicted
 F:349/Active site: Glu #status predicted

Query Match

Best Local Similarity 28.7%; Score 52.5; DB 2; Length 814;
 Pred. No. 77;
 Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

OY 3 IEGPTLRQWLARAGPNCIEGPTLRQWLA 31

Db 728 LKGPIC-QYRAAGSPSPRPGRPALIA 755

RESULT 18

T22892
 hypothetical protein F58B3.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T22892

R:Harris, B.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z19633
 A:Accession: T22892
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-214 <WLL>
 A:Cross-references: EMBL:Z73427; PID:CAA97797.1; GSPDB:GN00022; CESP:F58B3.1
 A:Experimental source: clone F58B3
 C:Genetics:
 A:Gene: CESP:F58B3.1
 A:Map position: 4
 A:Insertions: 68/1
 C:Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match

Best Local Similarity 28.4%; Score 52; DB 2; Length 214;
 Pred. No. 22;
 Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 1 GIEGPTLRQWLARAGPNCI 21

Db 186 GWSRPTIHWEGTTPGCGV 206

RESULT 19

T32999
 hypothetical protein F17E9.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T32999
 R:Moessner, J.
 submitted to the EMBL Data Library, February 1998
 A:Description: The sequence of C. elegans cosmid F17E9.
 A:Reference number: Z21262
 A:Accession: T32999
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-230 <NOE>
 A:Cross-references: EMBL:AF047656; PID:MAC05110.1; GSPDB:GN00022; CESP:F17E9.11
 A:Experimental source: strain Bristol N2; clone F17E9
 C:Genetics:
 A:Gene: CESP:F17E9.11
 A:Map position: 4
 C:Superfamily: Caenorhabditis elegans hypothetical protein F58B3.3

Query Match

Best Local Similarity 28.4%; Score 52; DB 2; Length 230;
 Pred. No. 24;
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 GIEGPTLRQWLARAGPNCI 21

Db 202 GGWKKPTIHWGHTTNGPCGV 222

RESULT 20

T35254
 conserved hypothetical protein SC5F2A.12c - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 18-Aug-2000
 C:Accession: T35254
 R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A
 submitted to the EMBL Data Library, April 1999
 A:Reference number: Z21573
 A:Accession: T35254
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-396 <OLI>
 A:Cross-references: EMBL:AL049587; PID:CA840679.1; GSPDB:GN00070; SC0EDB:SC5F2A.12c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SC0EDB:SC5F2A.12c

A:Residues: 1-281 <KAM>
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDD:BA09847.1; PID:di043633; PID:g5104188
A:Experimental source: strain KI
C:Genetics:
A:Gene: APE0867
C:Superfamily: Aeropyrum pernix hypothetical protein APE0867

Query Match	27.9%	Score 51	DB 2	Length 281
Best Local Similarity	34.4%	Pred. No. 39		
Matches 11: Conservative	7	Mismatches 6	Indels 8	Gaps 2
Oy	7	TUROWLARAAGPN-----GIEGPTLRQMLAAR	33	
		:::		
Db	12	SLRQMMRS---PNRYDIPQVDSPEYGVWLESR	40	

RESULT 26
D70601
UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) galU [similarity] - Mycobacter
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence,revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70601
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544 1998
A:Authors: Squares, R., Sultston, J.E., Taylor, K., Whitehead, S., Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70601
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-306 <COL>
A:Cross-references: GB:Z94752; GB:AL123456; NID:g3261731; PIDN:CAB08153.1; PID:g2052127
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: galU
C:Superfamily: Escherichia coli UTP--glucose-1-phosphate uridylyltransferase
;Keywords: nucleotidyltransferase

Query Match	27.9%	Score 51;	DB 2;	Length 306;
Best Local Similarity	69.2%	Pred. No. 43;		
Matches	9;	Conservative	1;	Mismatches 3; Indels 0; Caps 0;
Qy	5	GPILROWLAKRAG	17	
	::			
Db	290	GPDLKRWLVARLG	302	

RESULT 27
T29299 hypothetical protein C50F7.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T29299
R:Johnson, D.; Stellyes, L.
submitted to the EMBL data library, November 1995
A:Description: The sequence of C. elegans cosmid C50F7.
A:Reference number: Z20501
A:Accession: T29299
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-589 <TOH>
A:Cross-references: EMBL:U01557; PIDD:AAA83307.1; CESP:C50F7.2
C:Genetics:
A:Gene: CESP:C50F7.2
A:Introns: 12/2
A:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match	27.9%;	Score 51;	DB 2;	Length 589;
Best Local Similarity	42.9%;	Pred. No. 84;		
Matches	9;	Conservative	3;	Mismatches 9; Indels 0; Gaps 0

```
OY      4  EGPTRLRQWLARAGPNGIEGP 24
          | | : | | : | | : | |
Db      539 ESPSFQWIFGRPKPSGPAGP 559
```

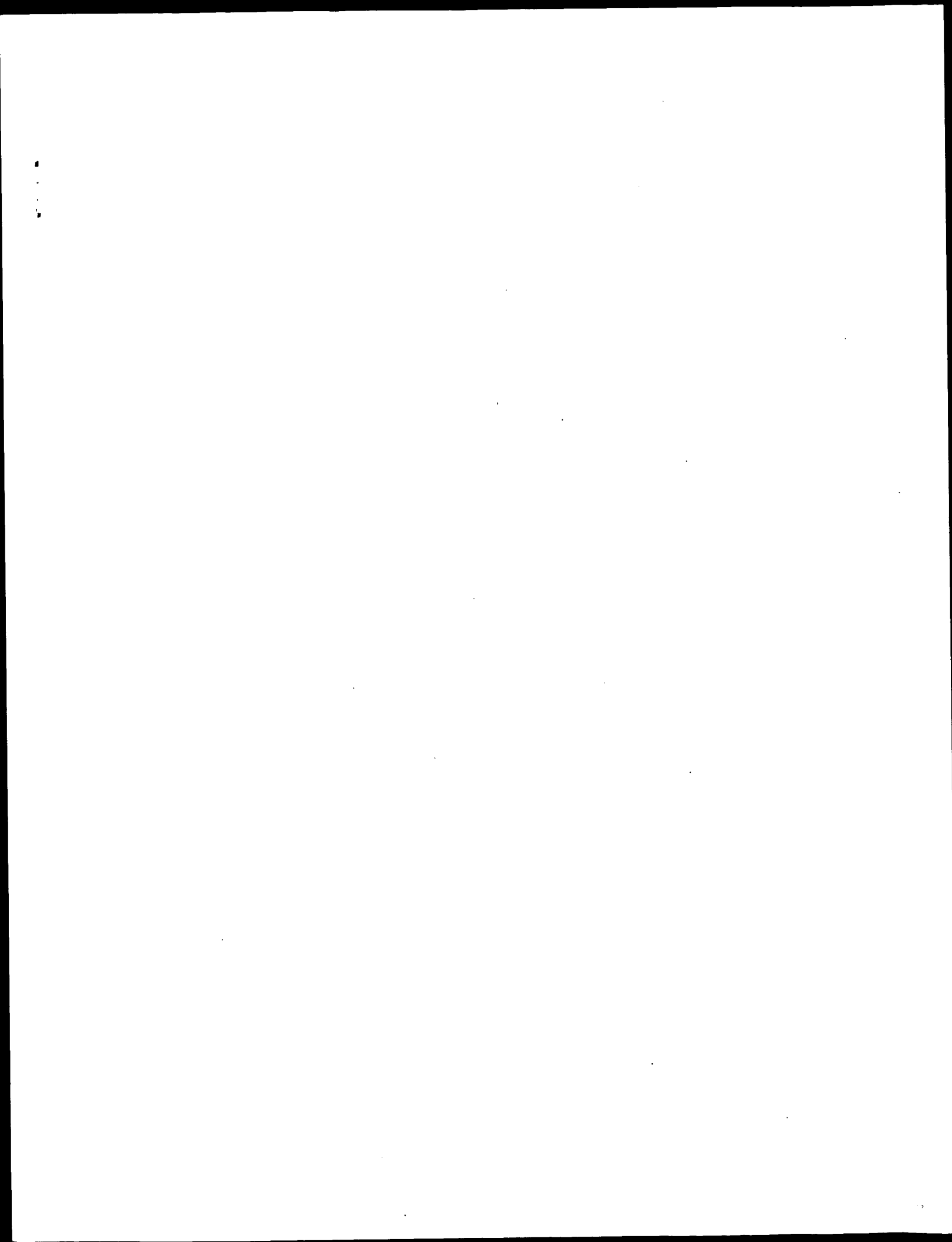
RESULT 28
C83221
Transport protein HasD PA3406 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence:revision 15-Sep-2000 #text-change 31-Dec-2000
C:Accession: C83221
R:Stevenson, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; L
: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
A:Reference number: AB2950, MUID:2043737
A:Accession: C83221
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <STO>
A:Cross-references: GB:AE004761; GB:AE004091; NID:g9949553; PIDN:AA606794.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: hasD, PA3406
A:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

	Query Match	27.9%	Score 51;	DB 2;	Length 600;
	Best Local Similarity	52.9%	Pred. No. 86;		
Matches	9; Conservative	3;	Mismatches	5;	Indels 0; Gaps 0;
QY	3 IEGETLRQWLAAKAGPN 19				
	: : : :				
Db	392 LDGADLRLQWSAALGPL 408				

RESULT 29
S04987
SIR5-binding protein sp105 - Pacific electric ray
C:Species: Torpedo californica (Pacific electric ray)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text-change 22-Jun-1999
C:Accession: S04987; S30070
R:Deutsch, I.J.; Garcia, A.M.; Lodish, H.F.
Biochem. J. 261, 155-166, 1989
A:Title: Primary structure of a novel 4-acetamido-4'-isochloroanostilbene-2,2'-disulph
A:Reference number: S04987; MUID: 89374082
A:Accession: S04987

	Query Match	27.9%	Score 51;	DB 1:	length 697;
	Best Local Similarity	42.1%;	Pred. No.	1e+02;	
Matches	8; Conservative	3;	Mismatches	8;	Indels 0; Gaps 0;
QY	11 WLAARAGPRTGTCPTLRW	29			
	: :				
Ddb	378 WLGLPSAANGSGGPLMKR	396			

RESULT 30
B95325
conserved hypothetical protein Sma0937 [imported] - *Sinorhizobium meliloti* (strain 102c)



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.06089 Seconds

(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838C-25

Perfect score: 183

Sequence: 1 GGIEGPTLRQWLAARAGPNEGPTLRQWLAARA 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	29.5	1366	1 CA21_HUMAN	P08123 homo sapien
2	53.5	29.2	266	2 SC02_HUMAN	O4819 homo sapien
3	53	29.0	524	1 VGGG_CHAV	P13180 chandipura
4	53	29.0	1372	1 CA21_MOUSE	O01149 mus musculu
5	52.5	28.7	814	1 AD15_HUMAN	Q13444 homo sapien
6	52	28.4	1372	1 CA21_RAT	P03466 rattus norv
7	51	27.9	696	1 SPI5_TORCA	P19965 torpedo cal
8	51	27.9	1838	1 CA15_HUMAN	P20908 homo sapien
9	50.5	27.6	904	1 DP01_MYCTU	O07700 mycobacteri
10	50	27.3	243	1 DERM_HUMAN	O99bhl homo sapien
11	50	27.3	246	1 TONB_PASHA	P72204 pasteurilla
12	50	27.3	410	1 ODBA_PSEPU	P09060 pseudomonas
13	50	27.3	415	1 TRMU_SCHPO	O13947 schizosacch
14	50	27.3	472	1 PSBC_SYNY3	P09193 synchocyst
15	50	27.3	1461	1 IE18_PPRVA	P33479 pseudorabie
16	50	27.3	1461	1 IE18_PPRVA	P11675 pseudorabie
17	49.5	27.0	333	1 CBBR_XANFL	P25545 xanthodacte
18	49.5	27.0	392	1 SERA_ECOLI	P11675 escherichia
19	49	26.8	368	1 ODBA_BACST	P21873 bacillus st
20	49	26.8	385	1 DIAC_HUMAN	O01459 homo sapien
21	49	26.8	735	1 CNG1_CHICK	O09805 gallus gall
22	49	26.8	911	1 CA1B_BOVIN	O28083 bos taurus
23	49	26.8	1669	1 CA14_MOUSE	P03463 mus musculu
24	49	26.8	2944	1 CA17_HUMAN	O03388 homo sapien
25	48.5	26.5	122	1 UROC_MOUSE	P81615 mus musculu
26	48.5	26.5	324	1 CCSI_CAEEL	P12114 caenorhabdi
27	48.5	26.5	562	1 SYR_AERPE	O9yft9 aeropyrum p
28	48	26.2	72	1 VXS_BP434	P11683 bacterioph
29	48	26.2	72	1 VXS_LAMB	P03699 bacterioph
30	48	26.2	270	1 YL6_VIBCH	O9yq28 vibrio chol
31	48	26.2	297	1 XERC_MYCLE	O9yq28 mycobacteri
32	48	26.2	298	1 TRPI_PSESK	P34818 pseudomonas
33	48	26.2	369	1 CA12_CHICK	P02460 gallus gall

ALIGNMENTS

```

RESULT 1
CA21_HUMAN          STANDARD:      PRT: 1366 AA.
ID                  CA21_HUMAN
AC P08123; P02464; Q9UEB6; Q9UPH0;
DT 01-AUG-1988 (Rel. 08, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2..
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN 1[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88058962; Pubmed=2824475;
RA de Wet W.J., Bernard M.P., Benson-Chanda V., Chu M.-L., Dickson L.A.,
RA Weil D., Ramirez F.;
RT "Organization of the human pro-alpha 2(I) collagen gene.";
RL J. Biol. Chem. 262:16032-16036(1987).
RN 1[2]
RP SEQUENCE FROM N.A.
RX Korkko J.M., Earley J.J., Ala-Korkko L., Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by CGE and DNA sequencing in
RT 14 patients with mild OI (Type I). Identification of common sequences
RT for null allele mutations.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN 1[3]
RP SEQUENCE OF 1-765 FROM N.A.
RX TISSUE-Placenta; Pubmed=3421913;
RA Kuivaniemi H., Tromp G., Chu M.-L., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 2(I)
RT chain of human type I procollagen. Comparison with the chicken gene
RT confirms unusual patterns of gene conservation.";
RL Biochem. J. 252:633-640(1988).
RN 1[4]
RP SEQUENCE OF 181-1366 FROM N.A.
RX Kalicki J., Wamsley P., Gibson A.;
RT Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN 1[5]
RP SEQUENCE OF 623-1366 FROM N.A.
RX MEDLINE=83178919; Pubmed=668791;
RA Bernard M.P., Myers J.C., Chu M.-L., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Structure of a cDNA for the pro alpha 2 chain of human type I
RT procollagen. Comparison with chick cDNA for pro alpha 2(I) identifies
RT structurally conserved features of the protein and the gene.";
RL Biochemistry 22:1139-1145(1983).
RN 1[6]
RP SEQUENCE OF 80-96.
RX TISSUE-Skin;
RX MEDLINE=71038625; Pubmed=5529814;
RX Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";

```

RL Biochemistry 9:4699-4706(1970).
 RN [17]
 RN SEQUENCE OF 417-447.
 RC TISSUE-SKIN;
 RX MEDLINE-75008198; PubMed-4412529;
 RA Pietrek P.P., Furthmayr H., Kuehn K.;
 RT "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit
 RT and pig-skin collagen.";
 RL Eur. J. Biochem. 47:257-261(1974).
 RN [18]
 RP SEQUENCE OF 145-198 FROM N.A.
 RX MEDLINE-88298792; PubMed-3403536;
 RA Kuivaniemi H., Sapol C., Tromp G., Sippola-Thiele M., Prockop D.J.;
 RT "A 19-base pair deletion in the pro-alpha 2(I) gene of type I
 RT procollagen that causes in-frame RNA splicing from exon 10 to exon 12
 RT in a proband with atypical osteogenesis imperfecta and in his
 RT asymptomatic mother";
 RL J. Biol. Chem. 263:11407-11413(1988).
 RN [9]
 RP SEQUENCE OF 960-1351 FROM N.A.
 RC TISSUE-SKIN;
 RX MEDLINE-90304220; PubMed-2364107;
 RA Maekelae J.K., Vuorio T., Vuorio E.;
 RT "Growth-dependent modulation of type I collagen production and mRNA
 RT levels in cultured human skin fibroblasts";
 RL Biochim. Biophys. Acta 1049:171-176(1990).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-91184577; PubMed-2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans.";
 RL FASEB J. 5:2052-2060(1991).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97255959; PubMed-9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrill-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [12]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-91374476; PubMed-1895312;
 RA Byers P.H., Wallis G.A., Willing M.C.;
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
 RL J. Med. Genet. 28:433-442(1991).
 RN [13]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97169389; PubMed-9016532;
 RA Dalgleish R.;
 RT "The human type I collagen mutation database.";
 RL Nucleic Acids Res. 25:181-187(1997).
 RN [14]
 RP VARIANT EDS-VII-A2.
 RX MEDLINE-88059013; PubMed-3680255;
 RA Wirtz M.K., Gnanville R.W., Steinmann B., Rao V.H., Hollister D.W.;
 RT "Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids
 RT comprising the N-telopeptide region of a pro-alpha 2(I) chain.";
 RL J. Biol. Chem. 262:16376-16385(1987).
 RN [15]
 RP SEQUENCE OF 1090-1107 FROM N.A., AND VARIANT OI-IV ARG-1102.
 RX MEDLINE-88227975; PubMed-2897363;
 RA Wenstrup R.J., Cohn D.H., Cohen T., Byers P.H.;
 RT "Arginine for glycine substitution in the triple-helical domain of
 RT the products of one alpha 2(I) collagen allele (COL1A2) produces the
 RT osteogenesis imperfecta type IV phenotype.";
 RL J. Biol. Chem. 263:7734-7740(1988).
 RN [16]
 RP VARIANT OI-II ASP-997.
 RX MEDLINE-89123407; PubMed-2914942;
 RA Baldwin C.T., Constantinou C., Dumars K.W., Prockop D.J.;
 RT "A single base mutation that converts glycine 907 of the alpha 2(I)
 RT chain of type I procollagen to aspartate in a lethal variant of
 RT osteogenesis imperfecta. The single amino acid substitution near the
 RT carboxyl terminus destabilizes the whole triple helix.";
 RL J. Biol. Chem. 264:3002-3006(1989).
 RN [17]
 RP VARIANT OI-II SER-955.
 RX MEDLINE-89380165; PubMed-2777764;
 RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
 RT "Characterization of point mutations in the collagen COL1A1 and
 RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
 RL J. Biol. Chem. 264:15809-15812(1989).
 RN [18]
 RP VARIANT OI-II CYS-877.
 RA Fectala A., Westerhausen A., Morris G.M., Rooney J.E., Prockop D.J.;
 RT "Two cysteine substitutions in the type I procollagen genes (COL1A1
 RT and COL1A2) that cause lethal osteogenesis imperfecta. The location
 RT of glycine substitutions does not in any simple way predict their
 RT effects on protein function or phenotype.";
 RL Am. J. Hum. Genet. 47:A216-A216(1990).
 RN [19]
 RP VARIANT EDS-VII-A2.
 RX MEDLINE-90368825; PubMed-2394758;
 RA Weil D., D'Alessio M., Ramirez F., Eyre D.R.;
 RT "Structural and functional characterization of a splicing mutation in
 RT the pro-alpha 2(I) collagen gene of an Ehlers-Danlos type VII
 RT patient.";
 RL J. Biol. Chem. 265:16007-16011(1990).
 RN [20]
 RP VARIANTS OI-IV VAL-676.
 RX MEDLINE-91291136; PubMed-2064612;
 RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
 RT valine substitution in osteogenesis imperfecta type IV. Detection of
 RT the mutation and prenatal diagnosis by a chemical cleavage method.";
 RL Biochem. J. 276:765-770(1991).
 RN [21]
 RP VARIANTS OI CYS-349 AND CYS-736.
 RX MEDLINE-9115889; PubMed-1990009;
 RA Wenstrup R.J., Shrago-Howe A.W., Lever L.W., Phillips C.L.;
 RT "Byers P.H., Cohn D.H.;
 RT "The effects of different cysteine for glycine substitutions within
 RT alpha 2(I) chains. Evidence of distinct structural domains within the
 RT type I collagen triple helix.";
 RL J. Biol. Chem. 266:2590-2594(1991).
 RN [22]
 RP VARIANT OI-II ARG-784.
 RX MEDLINE-91340689; PubMed-1874719;
 RA Tsuneyoshi T., Westerhausen A., Constantinou C.D., Prockop D.J.;
 RT "Substitutions for glycine alpha 1-637 and glycine alpha 2-694 of
 RT type I procollagen in lethal osteogenesis imperfecta. The
 RT conformational strain on the triple helix introduced by a glycine
 RT substitution can be transmitted along the helix.";
 RL J. Biol. Chem. 266:15608-15613(1991).
 RN [23]
 RP VARIANT OI-IV SER-751.
 RX MEDLINE-91271401; PubMed-2052622;
 RA Spock L.D., Constantinou C.D., Sereda L., Ganguj A., Riggs B.L.,
 RT Prockop D.J.;
 RT "Mutation in a gene for type I procollagen (COL1A2) in a woman with
 RT postmenopausal osteoporosis: evidence for phenotypic and genotypic
 RT overlap with mild osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5423-5427(1991).
 RN [24]
 RP VARIANT OI-II ARG-547.
 RX MEDLINE-93244832; PubMed-1284475;
 RA Bateman J.F., Moeller I., Hannagan M., Chan D., Cole W.G.;
 RT "Lethal perinatal osteogenesis imperfecta due to a type I collagen
 RT alpha 2(I) Gly to Arg substitution detected by chemical cleavage of
 RT an mRNA:CDNA sequence mismatch.";
 RL Hum. Mutat. 1:55-62(1992).
 RN [25]
 RP VARIANT OI-II ASP-670.
 RX MEDLINE-93054637; PubMed-1385413;

Query Match 29.5%; Score 54; DB 1; Length 1366;
Best Local Similarity 52.2%; Pred. No. 34;
Matches 12; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 GIEGPTLRQMLARAGPNCIEGP 24
DB 751 GVGPGTGPVGAAGPAGPNCPPGP 773

RESULT 2

SC02_HUMAN STANDARD; PRT; 266 AA.
ID SC02_HUMAN
AC 043819; Q9UK87;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SC02 protein homolog, mitochondrial precursor.
GN SC02.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Monocytes;
RS Smit L.J., Burton J.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS FIC LYS-140 AND PHE-225.
RX MEDLINE=20014747; PubMed=10545952;
RA Papadopoulos L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,
RA Sedlock J.E., Krishna S., Walker W., Selby J., Glerum D.M.,
RA Van Coster R., Lyon G., Scalsis E., Lebel R., Kaplan P., Shanske S.,
RA De Vivo D.C., Bonilla E., Hirano M., Dimauro S., Schon E.A.,
RT "Fatal infantile cardencephalomyopathy with COX deficiency and
mutations in SC02, a COX assembly gene."
RL Nat. Genet. 23:333-337(1999).
CC -1- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER
TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.
CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -1- TISSUE SPECIFICITY: UROGUTIOUS.
CC -1- DISEASE: DEFECTS IN SC02 ARE THE CAUSE OF FATAL INFANTILE
CARDENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS
CHARACTERIZED BY HYPERTRYPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND
GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX
ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX
DEFICIENCIES.
CC -1- SIMILARITY: BELONGS TO THE SC01/2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF177385; AAF05313.1; -;
DR EMBL: AL021683; CA16671.1; -;
DR MIM: 604272; -;
DR MIM: 604377; -;
DR MIM: 220110; -;
DR InterPro: IPR003782; SC01_Senc.
DR pfam: PF02630; SC01_Senc.1.
KW Mitochondrion; Transit peptide; Disease mutation; Polymorphism.
FT TRANSIT 1 41
FT CHAIN 42 266 MITOCHONDRION (POTENTIAL).
FT SC02 PROTEIN HOMOLOG.
FT VARIANT 20 20 R -> P (IN DBSNP:140523).
FT VARIANT 140 140 /FTID=VAR_011738.
FT VARIANT 140 140 E -> K (IN FIC).
FT VARIANT 225 225 /FTID=VAR_008874.
FT VARIANT S -> F (IN FIC).

FT SQ SEQUENCE 266 AA; 29810 MW; BC2F40E057329BF3 CRC64; /FTID=VAR_008875.

Query Match 29.2%; Score 53.5; DB 1; Length 266;
Best Local Similarity 33.3%; Pred. No. 7.8;
Matches 16; Conservative 2; Mismatches 9; Indels 21; Gaps 2;

QY 8 LRQMLARAGP-----NGIEGPTLR-----QMLARA 34
DB 33 LRSMILSRQPAETGGGQPGGPIRTLLITGLFAGAGLGAAMLALRA 80

RESULT 3

VGLG_CHAV STANDARD; PRT; 524 AA.
ID VGLG_CHAV
AC P13180;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Spike glycoprotein precursor.
GN G.
OS Chandipura virus (strain 1653514).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Rhabdoviridae; Vesiculovirus.
OX NCBI_TaxID=11273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299473; PubMed=2741347;
RA Masters P.S., Bhella R.S., Butcher M., Patel B., Ghosh H.P.,
RA Banerjee A.K.;
RT "Structure and expression of the glycoprotein gene of Chandipura
virus."
RL Virology 171:285-290(1989).
CC -1- FUNCTION: THIS PROTEIN FORMS SPIKES ON THE SURFACE OF THE VIRION.
IT IS RESPONSIBLE BOTH FOR THE BINDING OF THE VIRUS TO SUSCEPTIBLE
CC HOST CELLS AND FOR INDUCING THE UPTAKE OF THE VIRUS BY THE CELL.
CC THE INTERACTION BETWEEN THE INTERNAL COMPONENTS OF THE VIRION
CC AND THE PORTION OF THE GLYCOPROTEIN EXPOSED ON THE CYTOPLASMIC
CC FACE OF THE PLASMA MEMBRANE PROBABLY DIRECTS ENVELOPMENT AND
CC VIRUS BUDDING.
CC -1- SUBUNIT: TRIMERS IN THE ENDOPLASMIC RETICULUM.
CC -1- PTM: THIS PROTEIN IS MODIFIED BY THE COVALENT ADDITION OF PALMITIC
ACID VIA A THIOETHER LINKAGE TO A CYSTEINE. IT COULD BE EITHER ON
CC POSITION 479 OR 484.
CC -1- SIMILARITY: 39% IDENTITY TO THE G PROTEINS OF VSV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04350; AAA42916.1; -;
DR PIR: A32443; VGVNVC.
DR InterPro: IPR001903; Rhabd_glycop.
DR pfam: PF00974; Rhabd_glycop.2.
KW Transmembrane; Envelope protein; Glycoprotein; Lipoprotein; Palmitate;
KW Signal.
KW SIGNAL 1 21
FT CHAIN 22 524 SPIKE GLYCOPROTEIN.
FT DOMAIN 22 472 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 473 496
FT DOMAIN 497 524
FT DOMAIN 184 184 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 479 479 PALMITATE (POTENTIAL).
FT LIPID 484 484 PALMITATE (POTENTIAL).
SO SEQUENCE 524 AA; 58826 MW; A84AFOA5FFB73CD CRC64;

Query Match 29.0%; Score 53; DB 1; Length 524;
Best Local Similarity 37.0%; Pred. No. 17;

CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACTS WITH INTEGRIN ALPHA-V-BETA3, ENDOPHYLIN I AND
 CC SORTING NEXTIN 9. ENDOPHYLIN I AND SORTING NEXTIN 9 PREPREFERENTIALLY
 CC BIND THE PRECURSOR BUT NOT THE PROCESSED FORM OF ADAM15.
 CC SUGGESTING THAT THE INTERACTION OCCURS IN A SECRETORY PATHWAY
 CC COMPARTMENT PRIOR TO THE MEDIAL GOLGI (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED. OVEREXPRESSED IN
 CC ARTERIOSCLEROTIC LESIONS. CONSTITUTIVELY EXPRESSED IN CULTURED
 CC ENDOTHELIAL AND SMOOTH MUSCLE.
 CC -1- DOMAIN: THE CYTOPLASMIC DOMAIN INTERACTS WITH ENDOPHYLIN I AND
 CC SORTING NEXTIN 9 (BY SIMILARITY).
 CC -1- DOMAIN: DESINTEGRIN DOMAIN BINDS TO INTEGRIN ALPHA-V-BETA3.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U46005; AAC5112.1; -
 CC EMBL: U41767; AAC50404.1; -
 CC HSSP: P18619; 1FVL.
 CC MEROPS: M12.215; -.
 CC MIM: 605548; -.
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001818; Matrxin.
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; Repeplysin.
 CC InterPro: IPR000130; Zn_Mrpeptidase.
 CC Pfam: PF00200; disintegrin; 1.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC Pfam: PF01421; Repeplysin; 1.
 CC ProDom: PD000664; Disintegrin; 1.
 CC SMART: SM00050; DISIN; 1.
 CC SMART: SM00181; EGF; 1.
 CC PROSITE: PS00215; ADAM_MEROP; 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC PROSITE: PS00214; DISINTEGRIN_2; 1.
 CC PROSITE: PS00022; EGF_1; FALSE_NEG.
 CC PROSITE: PS01186; EGF_2; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 CC Transmembrane; EGF-like domain; SH3-binding.
 CC SIGNAL 1 17
 CC PROPEP 18 206
 CC CHAIN 207 814
 CC DOMAIN 207 696
 CC TRANSMEM 697 717
 CC DOMAIN 718 814
 CC DOMAIN 207 414
 CC DOMAIN 421 508
 CC DOMAIN 509 656
 CC DOMAIN 657 685
 CC SITE 766 772
 CC SITE 801 807
 CC SITE 179 179
 CC SITE 484 486
 CC METAL 348 348
 CC ACT_SITE 349 349
 CC METAL 352 352
 CC METAL 358 358
 CC DISULFID 323 409
 CC DISULFID 480 493

FT DISULFID 657 667 BY SIMILARITY.
 FT DISULFID 661 673 BY SIMILARITY.
 FT DISULFID 675 684 BY SIMILARITY.
 FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 606 606 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 714 714 G -> S (IN REF. 2).
 FT CONFLICT 791 791 A -> P (IN REF. 2).
 SQ SEQUENCE 814 AA; 87686 MW; DD2EC26CB1314576 CRC64;
 Query Match 28.7%; Score 52.5; DB 1; Length 814;
 Best Local Similarity 44.8%; Pred. No. 31;
 Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;
 Oy 3 IEPTLRQWLAARAGPNCIGPTLRQWLA 31
 Db 728 LKGPTC-QYRAQSGSPERPQPORALLA 755
 RESULT 6
 CA21_RAT STANDARD; PRT; 1372 AA.
 ID CA21_RAT
 AC P02466; O9R1E8;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 2(I) chain precursor.
 GN COL1A2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Guenther D., Seibold S., Marx M.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 86-98.
 RC TISSUE=Skin;
 RX MEDLINE=67162268; PubMed=5337886;
 RA Kang A.H., Bornstein P., Piez K.A.;
 RT "The amino acid sequence of peptides from the cross-linking region of
 RT rat skin collagen.";
 RL Biochemistry 6:788-795(1967).
 RN [3]
 RP SEQUENCE OF 99-102.
 RC TISSUE=Skin;
 RX MEDLINE=69206881; PubMed=5785232;
 RA Fietzek P.P., Piez K.A.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 2 chain of rat skin collagen.";
 RL Biochemistry 8:2129-2133(1969).
 RN [4]
 RP SEQUENCE OF 102-144.
 RC TISSUE=Skin;
 RX MEDLINE=73049496; PubMed=4636752;
 RA Fietzek P.P., Kell I., Kuehn K.;
 RT "The covalent structure of collagen. Amino acid sequence of the N-
 RT terminal region of alpha 2-CB4 from calf and rat skin collagen.";
 RL FEBS Lett. 26:66-68(1972).
 RN [5]
 RP SEQUENCE OF 423-452.
 RC TISSUE=Skin;
 RX MEDLINE=71115216; PubMed=5544653;
 RA Hieberger J.H., Kang A.H., Gross J.;
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
 RT peptides from chick and rat skin collagens.";
 RN Biochemistry 10:610-616(1971).
 RP [6]
 RC SEQUENCE OF 453-501.
 RP TISSUE=Skin;
 RC

RA MEDLINE=75059250; PubMed=4435743;
 RA Fietzek P.P., Kuehn K.;
 RT "The covalent structure of collagen: amino acid sequence of the N-
 RT terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5
 RT from calf skin collagen.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).
 RN [7]
 RP SEQUENCE OF 791-836.
 RC TISSUE=SKIN;
 RX MEDLINE=74055004; PubMed=4763308;
 RA Fietzek P.P., Kuehn K.;
 RT "The covalent structure of collagen: amino acid sequence of the N-
 RT terminal region of alpha 2-CB5 from rat skin collagen.";
 RL FBS Lett. 36:289-291(1973).
 RN [8]
 RP ORDER OF CNBR PEPTIDES.
 RX MEDLINE=70181852; PubMed=5443712;
 RA Vuust J., Lane J.M., Fietzek P.P., Miller E.J., Piez K.A.;
 RT "The order of the CNBR peptides from the alpha 2 chain of collagen.";
 RL Biochem. Biophys. Res. Commun. 38:703-708(1970).
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 CC (FIBRILLAR FORMING COLLAGEN).
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 CC HYDROXYAPATITE.
 CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF121217; AAD4175.1; -;
 CC PIR; A02867; CGRT25.
 CC InterPro: IPR000087; Collagen.
 CC DR InterPro: IPR000885; Fib.collagen_C.
 CC Pfam: PF01391; Collagen; 18
 CC ProDom; PD002078; Fib.collagen_C; 1.
 CC SMART; SM00038; COLF1; 1.
 CC KMW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Glycoprotein; Collagen; Signal.
 CC FT SIGNAL 1 24
 CC FT PROPEP 25 85
 CC FT CHAIN 86 1108
 CC FT PROPEP 1109 1372
 CC FT SITE 783 785
 CC FT SITE 828 830
 CC FT SITE 1011 1013
 CC FT MOD.RES 86 86
 CC FT MOD.RES 90 90
 CC FT CARBOHYD 1273 1273
 CC FT CONFLICT 132 132
 CC FT CONFLICT 137 137
 CC FT CONFLICT 145 422
 CC FT CONFLICT 431 432
 CC FT CONFLICT 494 494
 CC FT CONFLICT 497 497
 CC FT CONFLICT 502 790
 CC FT CONFLICT 825 825
 CC FT SEQUENCE 1372 AA; 129564 MW; B069371A8DE20A72 CRC64;
 Query Match 28.4%; Score 52; DB 1; Length 1372;
 Best Local Similarity 52.2%; Pred. No. 60;
 Matches 12; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 2 GIEGPTLRQWLARAGPNGIEGP 24
 DB 757 GIVGPTGPVGAAGSGPNGPPGP 779
 RESULT 7
 SP15_TORCA STANDARD; PRT; 696 AA.
 AC P19965;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE SITS-binding protein (SP105).
 OS Torpedo californica (Pacific electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squales; Hypnosquales; Pristigastera; Batoidae;
 OC Torpediniformes; Torpedinidae; Torpedo.
 CC NCBI_TaxID=7787;
 CC [1]
 CC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 CC MEDLINE=89374082; PubMed=2775201;
 CC Jentsch T.J., Garcia A.M., Lodish H.F.;
 CC "Primary structure of a novel
 CC 4-acetamido-4'-isothiocyanostilbene-2,2'-disulphonic acid
 CC (SITS)-binding membrane protein highly expressed in Torpedo
 CC californica electropex.";
 CC Biochem. J. 261:155-166(1989).
 CC -1- FUNCTION: THIS GLYCOPROTEIN IS PROBABLY NOT A FUNCTIONAL PART OF
 CC THE CHLORIDE CHANNEL.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- TISSUE SPECIFICITY: ELECTROPLAX TISSUE. BRAIN (200-FOLD LESS), AND
 CC HEART (500-FOLD LESS).
 CC -1- MISCELLANEOUS: BINDS 4-ACETAMIDO-4'-ISOTHIOCYANOSTILBENE-2,2'-DIS
 CC ULPHONIC ACID (SITS), AN INHIBITOR OF A VARIETY OF ANION TRANSPORT
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X16078; CAA34209.1; -;
 CC PIR; S04987; S04987.
 CC InterPro: IPR000322; Glyco_hydro_31.
 CC Pfam: PF01055; Glyco_hydro_31; 1.
 CC Transmembrane; Glycoprotein.
 CC FT INIT_MET 0 0
 CC FT DOMAIN 1 28
 CC FT TRANSMEM 29 49
 CC FT TRANSMEM 502 520
 CC FT TRANSMEM 541 561
 CC FT CARBOHYD 111 111
 CC FT CARBOHYD 133 133
 CC FT CARBOHYD 161 161
 CC FT CARBOHYD 385 385
 CC FT CARBOHYD 404 404
 CC FT CARBOHYD 469 469
 CC FT CARBOHYD 567 567
 CC FT SEQUENCE 696 AA; 78325 MW; 50C4BE98AFAFBE84 CRC64;
 Query Match 27.9%; Score 51; DB 1; Length 696;
 Best Local Similarity 42.1%; Pred. No. 41;
 Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
 QY 11 WLARAGPNGIEGPTLRQW 29
 DB 377 WLGLPSANGSGPPLMKW 395
 RESULT 8

CA15_HUMAN
ID CA15_HUMAN STANDARD: PRT: 1838 AA.
AC P20908;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(V) chain precursor.
GN COL5A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP MEDLINE=91302336; PubMed=2071595;
RX Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaei Y., Kato I.;
RA "Complete primary structure of human collagen alpha 1 (V) chain.";
RT J. Biol. Chem. 266:13124-13129(1991).
RN [2]
RP SEQUENCE OF 621-822.
RC TISSUE=Chorioamniotic membrane;
RX MEDLINE=89227189; PubMed=2496661;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of three cyanogen bromide-derived peptides from human alpha 1(V) collagen chain.";
RT Arch. Biochem. Biophys. 271:120-129(1989).
RN [3]
RP SEQUENCE OF 823-950, AND HEPARIN-BINDING.
RX MEDLINE=90366601; PubMed=2203476;
RA Yaei Y., Hashimoto K., Koltashni H., Takahara K., Ito M., Kato I.;
RT "Primary structure of the heparin-binding site of type V collagen.";
RT Biochim. Biophys. Acta 1035:139-145(1990).
RN [4]
RP SEQUENCE OF 556-571.
RC TISSUE=Placenta;
RX MEDLINE=92239022; PubMed=1571108;
RA Mann K.;
RT "Isolation of the alpha 3-chain of human type V collagen and characterization by partial sequencing.";
RT Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN [5]
RP SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.
RC TISSUE=Chorioamniotic membrane;
RX MEDLINE=94237164; PubMed=818482;
RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champlaud M.F., Boulillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT "Diversity in the processing events at the N-terminus of type-V collagen.";
RT Eur. J. Biochem. 221:987-995(1994).
RN [6]
RP VARIANT ED51 SER-1639.
RX MEDLINE=97195540; PubMed=9042913;
RA de Paape A., Nuytink L., Hausser I., Anton-Lamprecht I., Naeyaert J.-M.;
RT "Mutations in the COL5A1 gene are causal in the Ehlers-Danlos syndromes I and II.";
RT Am. J. Hum. Genet. 60:547-554(1997).
RN [7]
RP FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE COMPONENT OF NEARLY UBQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
CC -1 SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND ONE ALPHA 3(V) CHAINS IN PLACENTA.
CC -1 PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1 PTM: 40% OF TYROSINES IN THE PRO-ALPHA 1(V) CHAIN ARE SULFATED.
CC -1 DISEASE: DEFECTS IN COL5A1 ARE A CAUSE OF EHLERS-DANLOS SYNDROME, TYPE I (EDS1), A DISEASE CHARACTERIZED BY LOOSE-JOINTEDNESS AND FRAGILE, VELVETY, STRETCHABLE, BRUISEABLE SKIN THAT HEALS WITH PECULIAR 'CIGARETTE-PAPER' SCARS.

CC -1 SIMILARITY: HIGH, TO ALPHA 3(V) AND ALPHA 1(XI) CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D90279; BA1423.1; -
CC PIR: S03978; S03978.
CC PIR: S11303; S11303.
CC PIR: S16024; S16024.
CC MIM: 120215; -
CC MIM: 130000; -
CC MIM: 130010; -
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR000885; Fib_collagen_C.
CC InterPro: IPR001791; Laminin_G.
CC InterPro: IPR003129; TSPN.
CC Pfam: PF01391; Collagen_18.
CC Pfam: PF02210; TSPN_1.
CC ProDom: PD002078; Fib_collagen_C_1.
CC SMART: SM00038; COLF1; 1.
CC SMART: SM00282; LamG; 1.
CC SMART: SM00210; TSPN; 1.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Collagen; Signal; Heparin-binding; Sulfation; Disease mutation.
CC CHAIN 1 37
CC SIGNAL 38 1605
CC FT 38 443 COLLAGEN ALPHA 1(V) CHAIN.
CC FT 444 558 NONHELMICAL REGION.
CC FT 559 1570 INTERRUPTED COLLAGENOUS REGION.
CC FT 1571 1605 TRIPLE-HELICAL REGION.
CC FT 1606 1858 NONHELMICAL REGION.
CC FT 1858 1858 CARBOXYL-TERMINAL PROPEPTIDE.
CC FT 570 570 HYDROXYLATION.
CC FT 576 576 HYDROXYLATION.
CC FT 621 621 HYDROXYLATION.
CC FT 627 627 HYDROXYLATION.
CC FT 639 639 HYDROXYLATION.
CC FT 642 642 HYDROXYLATION.
CC FT 648 648 HYDROXYLATION.
CC FT 654 654 HYDROXYLATION.
CC FT 657 657 HYDROXYLATION.
CC FT 675 675 HYDROXYLATION.
CC FT 678 678 HYDROXYLATION.
CC FT 680 680 HYDROXYLATION.
CC FT 686 686 HYDROXYLATION.
CC FT 690 690 HYDROXYLATION.
CC FT 696 696 HYDROXYLATION.
CC FT 705 705 HYDROXYLATION.
CC FT 708 708 HYDROXYLATION.
CC FT 717 717 HYDROXYLATION.
CC FT 720 720 HYDROXYLATION.
CC FT 726 726 HYDROXYLATION.
CC FT 732 732 HYDROXYLATION.
CC FT 744 744 HYDROXYLATION.
CC FT 750 750 HYDROXYLATION.
CC FT 756 756 HYDROXYLATION.
CC FT 762 762 HYDROXYLATION.
CC FT 765 765 HYDROXYLATION.
CC FT 771 771 HYDROXYLATION.
CC FT 774 774 HYDROXYLATION.
CC FT 780 780 HYDROXYLATION.
CC FT 789 789 HYDROXYLATION.
CC FT 795 795 HYDROXYLATION.
CC FT 804 804 HYDROXYLATION.
CC FT 807 807 HYDROXYLATION.
CC FT 810 810 HYDROXYLATION.
CC FT 816 816 HYDROXYLATION.
CC FT 819 819 HYDROXYLATION.

Db 318 GGALPAGTVROWLAHAGDGRAGLTV 344

RESULT 10

DEFL_HUMAN

ID DEFL_HUMAN

STANDARD:

PRT; 243 AA.

DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Peptide deformylase, mitochondrial precursor (EC 3.5.1.88) (PDF)

DE (Polypeptide deformylase).

CN PDPLA OR PDF.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_TaxID=9606;

RN [1]

RX MEDLINE=20514156; PubMed=11060042;

RA Gignone C., Sereto A., Piere M., Boisson B., Meinel T.;

RT "Identification of eukaryotic peptide deformylases reveals

RT universality of N-terminal protein processing mechanisms.";

RL EMBO J. 19:5916-5929(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Lonetto M.A., Zhu Y., Li X., Southan C.;

RT "A human homolog of bacterial peptide deformylases."

RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

CC -! FUNCTION: Removes the formyl group from the N-terminal Met of

CC newly synthesized proteins (By similarity).

CC -! CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +

CC methionyl peptide.

CC -! COFACTOR: Binds 1 iron(II) ion (By similarity).

CC -! SUBCELLULAR LOCATION: Mitochondrial (Potential).

CC -! TISSUE SPECIFICITY: Ubiquitous.

CC -! SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF239156; AAK3968.1; -

DR EMBL; AF232879; AAK15624.1; -

DR InterPro; IPR000181; Pep.deformylase.

DR Pfam; PF01327; Pep.deformylase; 1.

DR Prodom; PD003844; Pep.deformylase; 1.

DR Protein biosynthesis; Hydrolyase; Iron; Mitochondrion; Transit peptide.

FT TRANSIT 1 ? 243 MITOCHONDRION (POTENTIAL).

FT CHAIN 1 ? 243 PEPTIDE DEFORMYLASE.

FT METAL 172 172 IRON (BY SIMILARITY).

FT METAL 214 214 IRON (BY SIMILARITY).

FT ACT_SITE 215 215 BY SIMILARITY.

FT METAL 218 218 IRON (BY SIMILARITY).

SQ SEQUENCE 243 AA; 27013 MW; B15A3456F0F8D689 CRC64;

Db 31 SSTAADGVEGALRR 46

Query Match 27.3%; Score 50; DB 1; Length 243;

Best Local Similarity 50.0%; Pred. No. 19;

Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 13 AARAGNGIEGPTLR 28

DB 31 SSTAADGVEGALRR 46

QY 13 AARAGNGIEGPTLR 28

DB 31 SSTAADGVEGALRR 46

QY 13 AARAGNGIEGPTLR 28

DB 31 SSTAADGVEGALRR 46

QY 13 AARAGNGIEGPTLR 28

DB 31 SSTAADGVEGALRR 46

QY 13 AARAGNGIEGPTLR 28

DB 31 SSTAADGVEGALRR 46

QY 13 AARAGNGIEGPTLR 28

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE TonB protein.

GN TonB.

OS Pasteurella haemolytica.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Mannheimia.

NCBI_TaxID=75985;

RN [1]

RP SEQUENCE FROM N.A.

RA Graham M.R., Lo R.Y.C.;

RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.

CC -! FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT

CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO

CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO

CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-

CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE

CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER

CC MEMBRANE PROTEINS (BY SIMILARITY).

CC -! SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC

CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE

CC PERIPLASM (BY SIMILARITY).

CC -! SIMILARITY: BELONGS TO THE TONB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U6265; AAB09530.1; -

DR KW Transport; Protein transport; Inner membrane; Periplasmic;

FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 8 28 SIGNAL-ANCHOR (POTENTIAL).

FT DOMAIN 29 246 PERIPLASMIC (POTENTIAL).

SQ SEQUENCE 246 AA; 27785 MW; C9582F619FCBA5B5 CRC64;

Query Match 27.3%; Score 50; DB 1; Length 246;

Best Local Similarity 47.4%; Pred. No. 19;

Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

DB 157 GPEIKGIVAKAIPMAEG 175

QY 5 GPTLRQWLAARAGNGTIG 23

RT branched-chain-oxoacid and pyruvate dehydrogenases.";
 RL Eur. J. Biochem. 176:311-317(1988).
 RP [2]
 RC SEQUENCE OF 1-17 FROM N.A.
 RA MEDLINE=91008935; PubMed=2211503;
 RX Madhusudan K.T., Huang G., Burns G., Sokatch J.R.;
 RT "transcriptional analysis of the promoter region of the *Pseudomonas*
 RT *putida* branched-chain keto acid dehydrogenase operon.";
 RL J. Bacteriol. 172:5655-5663(1990).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RA MEDLINE=99356017; PubMed=10426958;
 RX Aeversson A., Seger K., Turley S., Sokatch J.R., Hol W.G.J.;
 RT "Crystal structure of 2-oxoisovalerate and dehydrogenase and the
 RT architecture of 2-oxo acid dehydrogenase multienzyme complexes.";
 RL Nat. Struct. Biol. 6:785-792(1999).
 CC -1- FUNCTION: THE BRANCHED-CHAIN ALPHA-KETO DEHYDROGENASE COMPLEX
 CC CATALYZES THE OVERALL CONVERSION OF ALPHA-KETO ACIDS TO ACYL-COA
 CC AND CO(2). IT CONTAINS MULTIPLE COPIES OF 3 ENZYMATIC COMPONENTS:
 CC BRANCHED-CHAIN ALPHA-KETO ACID DECARBOXYLASE (E1), LIPOAMIDE
 CC ACYLTRANSFERASE (E2) AND LIPOAMIDE DEHYDROGENASE (E3).
 CC -1- CATALYTIC ACTIVITY: 3-methyl-2-oxobutanoate + lipamide = S-(2-
 CC methylpropanoyl)dihydroliipoamide + CO(2).
 CC -1- COPFACTOR: THIAMINE PYROPHOSPHATE.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M57613; AAA65614.1; -
 DR PIR: S01317; DEPSXA.
 DR PDB: 10S0; 18-AUG-99.
 DR InterPro: IPR001017; EL.dh.
 DR Pfam: PF00676; EL_dehydrocyt.1
 DR Oxidoreductase; Flavoprotein; Thiamine pyrophosphate; 3D-structure.
 KW SEQUENCE 410 AA; 45268 MW; 0C998460CCF89CF4 CRC64;
 SQ
 Query Match 27.3%; Score 50; DB 1; Length 410;
 Best Local Similarity 53.3%; Pred. NO. 32;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 OY 5 GPTLRQWLAARAGPN 19
 DB 298 GPTLRQWLAARAGPN 312
 ID TRMU_SCHPO STRAND: PRT; 415 AA.
 AC 013947;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Probable tRNA (5-methylaminomethyl-2-thiouridyate)-methyltransferase
 DE (EC 2.1.1.61).
 GN SPAC23H4.04.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Schizosaccharomycetes; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
 RL Submitted (SEP-1997) to the EMBL/Genbank/DBS databases.
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-

CC homocysteine + tRNA containing 5-methylaminomethyl-2-
 CC thouridyate.
 CC -1- SIMILARITY: BELONGS TO THE TRMU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z98977; CAB11659.1; -
 DR InterPro: IPR004135; tRNA_Me.trans.
 DR Pfam: PF03054; tRNA_Me.trans.1.
 KW Transferase; Methyltransferase; tRNA processing.
 SQ SEQUENCE 415 AA; 47626 MW; D2604335B7A935F CRC64;
 Query Match 27.3%; Score 50; DB 1; Length 415;
 Best Local Similarity 37.0%; Pred. NO. 33;
 Matches 10; Conservative 4; Mismatches 11; Indels 2; Gaps 1;
 OY 3 IEGPTLRQWLAARAGNGIEGPTLRW 29
 DB 58 VEGVFMKRMLEDSDASGSC-PAERDW 82
 ID PSBC_SYNY3 STRAND: PRT; 472 AA.
 AC P09193; P73749;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Photosystem II 44 kDa reaction center protein (P6 protein) (CP43).
 GN PSBC OR SL0851.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chisholm D., Williams J.G.K.;
 RT "Nucleotide sequence of psbC, the gene encoding the CP-43 chlorophyll
 RT a-binding protein of photosystem II, in the cyanobacterium
 RT Synechocystis 6803.";
 RL Plant Mol. Biol. 10:293-301(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 RN [3]
 RP SEQUENCE OF 1-300 FROM N.A.
 RX MEDLINE=88211542; PubMed=3130247;
 RA Delzalkins V.A., Bogorad L.;
 RT "Molecular analysis of a mutant defective in photosynthetic oxygen
 RT evolution and isolation of a complementing clone by a novel screening
 RT procedure.";
 RL EMBO J. 7:333-338(1988).
 CC -1- FUNCTION: THE 43 kDa PROTEIN (P6) IS A COMPONENT OF THE CORE OF
 CC PHOTOSYSTEM II. IT IS A CHLOROPHYLL BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; CELLULAR
 CC THYLAKOID MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M21538; AAA85378.1; -
 DR EMBL: D90909; BAA17799.1; -
 DR EMBL: X07018; CAA30071.1; -
 DR PIR: S06469; S06469.
 DR PIR: S02380; S02380.
 DR InterPro: IPR000932; PSII.
 DR Pfam: PF00421; PSII; 1.
 KM Photosynthesis II: Thylakoid; Chlorophyll;
 KW Transmembrane: Complete proteome.
 FT TRANSMEM 56 75 POTENTIAL.
 FT TRANSMEM 107 129 POTENTIAL.
 FT TRANSMEM 160 182 POTENTIAL.
 FT TRANSMEM 202 224 POTENTIAL.
 FT TRANSMEM 237 259 POTENTIAL.
 FT TRANSMEM 269 291 POTENTIAL.
 FT TRANSMEM 423 445 POTENTIAL.
 FT CONFLICT 54 54 R -> A (IN REF. 2).
 FT CONFLICT 67 67 T -> N (IN REF. 3).
 FT CONFLICT 162 162 Y -> I (IN REF. 3).
 SQ SEQUENCE 472 AA; 51760 MW; D94D9FE73F66192D CRC64;

Query Match 27.3%; Score 50; DB 1; Length 472;
 Best Local Similarity 35.0%; Pred. No. 37;
 Matches 14; Conservative 3; Mismatches 13; Indels 10; Gaps 2;

OY 5 GPTLR-----OMLAARAGPNGIEPTLR-----MLARA 34
 DB 352 GETMRWDFRGWLEPLRGPNGLDLDKLNIDIQPQVRRRA 391

RESULT 15
 IE18.PRVA

ID IE18.PRVA STANDARD; PRT; 1446 AA.

AC P33479;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Immediate-early protein IE180.

GN IE.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirinae.
 OX NCBI_TaxID=33703;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-91021039; PubMed-2171211;

RT Vicek C., Kozmuk Z., Paces V., Schirm S., Schwyzer M.;

RT "Pseudorabies virus immediate-early gene overlaps with an oppositely

RT oriented open reading frame: characterization of their promoter and

RT enhancer regions.";

RT Virology 179:365-377(1990).

CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE

CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING

CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.

CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.

CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF

CC PHOSPHORYLATION.

CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M34651; AAA47470.1; -
 DR PIR: A45344; A45344.
 KW Early protein; Transcription regulation; Trans-acting factor;
 KW DNA-binding; Phosphorylation; Nuclear protein.
 FT DOMAIN 347 354 POLY-SER.
 FT DOMAIN 379 397 POLY-SER.
 SQ SEQUENCE 1446 AA; 148640 MW; 81F43A3DE3DDA068 CRC64;

Query Match 27.3%; Score 50; DB 1; Length 1446;
 Best Local Similarity 42.9%; Pred. No. 1,1e+02;
 Matches 12; Conservative 2; Mismatches 12; Indels 2; Gaps 1;

OY 2 GIEGPTL--RQWLARAGPNGIEPTLR 27
 DB 179 GSPGSAAPRRWSPARQDPVGEPPAPAR 206

RESULT 16
 IE18.PRVI

ID IE18.PRVI STANDARD; PRT; 1461 AA.

AC P11675;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Immediate-early protein IE180.

GN IE.
 OS Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirinae.

OX NCBI_TaxID=31523;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-89315207; PubMed-2546124;

RA Cheung A.K.;

RT "DNA nucleotide sequence analysis of the immediate-early gene of

RT pseudorabies virus.";

RT Nucleic Acids Res. 17:4637-4646(1989).

RL [2]

RP REVISIONS.

RA Cheung A.K.;

RL Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE

CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING

CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.

CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.

CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF

CC PHOSPHORYLATION.

CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X15120; CAA33214.1; -

DR PIR: S04713; EDBE1F.

RW Early protein; Transcription regulation; Trans-acting factor;

FW DNA-binding; Phosphorylation; Nuclear protein.

FT DOMAIN 390 405 POLY-SER.

FT DOMAIN 958 966 POLY-SER.

SQ SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 27.3%; Score 50; DB 1; Length 1461;
 Best Local Similarity 42.9%; Pred. No. 1,1e+02;
 Matches 12; Conservative 2; Mismatches 12; Indels 2; Gaps 1;
 OY 2 GIEGPTL--RQWLARAGPNGIEPTLR 27
 DB 187 GSPGSAAPRRWSPARQDPVGEPPAPAR 214

```

RESULT 17
CBBR_XANFL STANDARD: PRT: 333 AA.
ID CBBR_XANFL STANDARD: PRT: 333 AA.
AC P25545;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Rubisco operon transcriptional regulator.
GN CBBR OR CFPO.
OS Xanthobacter flavus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hypnomicrobium group; Xanthobacter.
OX CBBL_TaxID=281;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H4-14;
RX MEDLINE=94012468; PubMed=8407781;
RA van den Bergh E., Dijkhuizen L., Meijer W.G.;
RT "Cbdr, a lysr-type transcriptional activator, is required for
RT expression of the autotrophic CO2 fixation enzymes of Xanthobacter
RT flavus." ;
RL J. Bacteriol. 175:6097-6104(1993).
RN [2]
RP SEQUENCE OF 1-150 FROM N.A.
RC STRAIN=H4-14;
RX MEDLINE=91172133; PubMed=190916;
RA Meijer W.G., Arnborg A.C., Enequist H.G., Terpstra P., Lidstrom M.E.,
RA Dijkhuizen L.;
RT "Identification and organization of carbon dioxide fixation genes in
RT Xanthobacter flavus H4-14." ;
RL Mol. Gen. Genet. 225:320-330(1991).
CC -I- FUNCTION: TRANSCRIPTIONAL ACTIVATOR FOR THE CBB OPERON (CBBLSXFP)
CC FOR RUBISCO AND OTHER CALVIN CYCLE GENES. BINDS SPECIFICALLY TO
CC TWO BINDING SITES IN THE CBBR-CBBL INTERGENIC REGION.
CC -I- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC CC
DR EMBL; Z22705; CAA80406.1; .
DR EMBL; X17252; ?; NOT_ANNOTATED_CDS.
DR PIR; S13578; S13578.
DR InterPro; IPR000847; HTH_LysR.
DR Pfam; PF00126; HTH.1; 1.
DR PROSITE; PS00044; HTH_LYSR_FAMILY; 1.
KM Transcription regulation; Activator; DNA-binding.
KT DN_BIND 22 41 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 333 AA; 36003 MW; 9B375B4FB2DIEE73 CRC64;
Query March 27.0%; Score 49.5; DB 1; Length 333;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;
QY 3 IEQ-PTLRQLAARA 16
DB 264 VEGLEPVVRQMLAVRA 278
SETA_ECOLI STANDARD: PRT: 392 AA.
AC P31675; P75639;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```


FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 366 386 POTENTIAL.
FT CONFLICT 56 56 T -> A (IN REF. 1).
FT CONFLICT 128 128 E -> A (IN REF. 1).
SQ SEQUENCE 392 AA; 42771 MW; 8DE9728BBE3D4460 CRC64;

Query Match 27.0%; Score 49.5; DB 1; Length 392;
Best Local Similarity 28.9%; Pred. No. 36;
Matches 13; Conservative 6; Mismatches 15; Indels 11; Gaps 1;

OY 1 GGIEGPTLROWLARAAGP-----NGIEGPTLROWLARA 34
DB 29 GAIQAPTLISLISREVGAQPFMIGLYTNALAGVSLMAKRS 73

RESULT 19
ODPA_BACST STANDARD; PRT; 368 AA.
AC P21873;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Pyruvate dehydrogenase E1 component, alpha subunit (EC 1.2.4.1).
GN PDHA.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=NCA 1503;
RX MEDLINE=90345939; PubMed=2200674;
RX Hawkins C.F., Borges A., Perham R.N.;
RT "Cloning and sequence analysis of the genes encoding the alpha and beta subunits of the E1 component of the pyruvate dehydrogenase multienzyme complex of Bacillus stearothermophilus.";
RL Eur. J. Biochem. 191:337-346(1990).
CC -1- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE COPIES OF THREE ENZYMAIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1), DIHYDROLIPONAMIDE ACETYLTRANSFERASE (E2) & LIPONAMIDE DEHYDROGENASE (E3).
CC -1- CATALYTIC ACTIVITY: Pyruvate + liponamide = S-acetylhydroliponamide + CO(2).
CC -1- COFACTOR: THIAMINE PYRROPHOSPHATE.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: X53560; CAA37628.1; -
DR PIR: S10798; DEBSPF.
DR HSSP: P09060; 10S0.
DR InterPro: IPR001017; EL.dh.
DR Pfam: PF00676; EL.dehydrog.1.
KW Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
FT Phosphorylation.
FT INIT_MET 0
FT MOD_RES 283 283 PHOSPHORYLATION.
SQ SEQUENCE 368 AA; 41338 MW; 46199FEF69EEA662 CRC64;

Query Match 26.8%; Score 49; DB 1; Length 368;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 12; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

OY 4 EGPTRWLARAAGPNCIEG--PT 25
DB 256 EGPTRWLARAAGPNCIEG--PT 279

RESULT 20
DIAC_HUMAN STANDARD; PRT; 385 AA.
AC Q01459;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Di-N-acetylchitinase precursor (EC 3.2.1.-).
GN CHS OR CTB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=92406917; PubMed=1527079;
RX Fisher K.J., Aronson N.N. Jr.;
RT "Cloning and expression of the cDNA sequence encoding the lysosomal glycosidase di-N-acetylchitinase.";
RL J. Biol. Chem. 267:19607-19616(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Liu B., Aronson N.N. Jr.;
RT "Structure of the human gene for lysosomal di-N-acetylchitinase.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE DEGRADATION OF ASPARAGINE-LINKED GLYCOPROTEINS. HYDROLYZE OF N-ACETYL-BETA-D-GLUCOSAMINE (1-4)-N-ACETYLGLUCOSAMINE CHITOBIOSIDE CORE FROM THE REDUCING END OF THE BOND, IT REQUIRES PRIOR CLEAVAGE BY GLYCOSYLASPARAGINASE.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO FAMILY 18 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: M95767; AAA35684.1; -
DR EMBL: AF085706; AAC35852.1; JOINED.
DR EMBL: AF085700; AAC35852.1; JOINED.
DR EMBL: AF085701; AAC35852.1; JOINED.
DR EMBL: AF085702; AAC35852.1; JOINED.
DR EMBL: AF085703; AAC35852.1; JOINED.
DR EMBL: AF085704; AAC35852.1; JOINED.
DR EMBL: AF085705; AAC35852.1; JOINED.
DR PIR: A44102; A44102.
DR PIR: S27959; S27959.
DR MIM: 600873; -
DR InterPro: IPR001579; Chitinase_2.
DR PROSITE: PS01095; CHITINASE_18; 1.
KW Hydrolyase; Glycosidase; Signal; Lysosome; Glycoprotein.
FT SIGNAL 1 38
FT CHAIN 39 385
FT ACT_SITE 143 143 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 193 193 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 385 AA; 43759 MW; 0A9D14C8B26B52DE CRC64;

Query Match 26.8%; Score 49; DB 1; Length 385;
Best Local Similarity 37.9%; Pred. No. 40;
Matches 11; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

OY 6 PTLRWLARAAGPNCIEGPTLROWLARA 34
DB 4 PTLRWLARAAGPNCIEGPTLROWLARA 32

RESULT 21
 CNG1_CHICK STANDARD; PRT; 735 AA.
 AC 090805;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cyclic nucleotide gated channel, cone photoreceptor, alpha subunit
 DE (CNG channel 1) (CNG-1).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBL_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93264082; PubMed=7684234;
 RA Boenigk W., Altenhofen W., Mueller F., Dose A., Illing M.,
 RA Molday R.S., Kaupp U.B.;
 RT "Rod and cone photoreceptor cells express distinct genes for
 RT GMP-gated channels."
 RL Neuron 10:865-877(1993).
 CC -1- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
 CC COUPLED CASCADE USING GMP AS SECOND MESSENGER. THIS PROTEIN CAN
 CC BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO A OPENING OF THE CATION
 CC CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF CONE
 CC PHOTORECEPTORS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X89598; CAA61757.1;
 DR InterPro: IPR000636; Cation_chan_non_119.
 DR Pfam: PF00027; CNMP_binding_1.
 DR Pfam: PF00520; ion_trans_1.
 DR SMART: SM00100; CNMP: 1.
 DR PROSITE: PS00888; CNMP_BINDING_1; 1.
 DR PROSITE: PS00889; CNMP_BINDING_2; 1.
 DR PROSITE: PS50042; CNMP_BINDING_3; 1.
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane; Vision;
 KW Multigene family.
 FT DOMAIN 1 210 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 211 230 H1 (POTENTIAL).
 FT DOMAIN 231 243 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 244 262 H2 (POTENTIAL).
 FT DOMAIN 263 286 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 287 306 H3 (POTENTIAL).
 FT DOMAIN 307 344 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 345 367 H4 (POTENTIAL).
 FT DOMAIN 368 419 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 420 439 H5 (POTENTIAL).
 FT DOMAIN 440 523 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 524 544 H6 (POTENTIAL).
 FT DOMAIN 545 735 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 532 654 CAMP (BY SIMILARITY).
 FT BINDING 591 606 CAMP (POTENTIAL).
 FT BINDING 606 606 CAMP (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 735 AA; 85031 MW; A67ADFD942CECE CRC64;

Query Match
 Best Local Similarity

26.8%; Score 49; DB 1; Length 735;
 34.1%; Pred. No. 76;

Matches 14; Conservative 2; Mismatches 11; Indels 14; Gaps 2;
 QY 3 IEPTL-----RQMLARAGPNCIEGPTLRQMLAR 33
 DB 103 IRGPELVSSRSQSNIRSLGIREPGGVNCP----WPLAR 139
 RESULT 22
 CALB_BOVIN STANDARD; PRT; 911 AA.
 ID CALB_BOVIN
 AC 028083;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Collagen alpha 1(XI) chain (Fragment).
 GN COL1A1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NC NCBL_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Smooth muscle;
 RX MEDLINE=92078200; PubMed=1744123;
 RA Brown K.E., Lawrence R., Sonenshein G.E.;
 RT "Concerted modulation of alpha 1(XI) and alpha 2(V) collagen mRNAs in
 RT bovine vascular smooth muscle cells."
 RL J. Biol. Chem. 265:23268-23273(1991).
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 1(XI) IS A POST-TRANSLATIONAL
 CC MODIFICATION OF ALPHA 1(XI). ALPHA 1(XI) CAN ALSO BE FOUND INSTEAD
 CC OF ALPHA 3(XI)-1(XI) (BY SIMILARITY).
 CC -1- PFM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.
 CC -1- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M82977; AAA30369.1;
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF01391; Collagen; 11.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen.
 FT NON_TER 1 1 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
 FT PROPEP <1 278 COLLAGEN ALPHA 1(XI) CHAIN.
 FT CHAIN 279 >911 NONHELICAL REGION.
 FT DOMAIN 187 186 TRIPLE-HELICAL REGION (INTERRUPTED).
 FT DOMAIN 187 275 TRIPLE-HELICAL REGION (INTERRUPTED).
 FT DOMAIN 276 278 SHORT NONHELICAL SEGMENT.
 FT DOMAIN 279 295 TELOPEPTIDE.
 FT DOMAIN 296 >911 TRIPLE-HELICAL REGION.
 FT SITE 379 379 CROSSLINKING.
 FT NON_TER 911 911
 SQ SEQUENCE 911 AA; 89259 MW; C05C4B350749CFC CRC64;

Query Match
 Best Local Similarity
 Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
 QY 3 IEPTLRQMLARAGPNCIEGPT 25
 DB 207 IEPPGAGPAGLMGPPDLGPT 229

RESULT 23
 CA14_MOUSE STANDARD: PRT: 1669 AA.
 ID AC CA14_MOUSE
 AC P02463;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN COL4A1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89197932; PubMed=2703490;
 RT Muthukumar G., Blumberg B., Kurkinen M.;
 RT "The complete primary structure for the alpha 1-chain of mouse
 RT collagen IV. Differential evolution of collagen IV domains.";
 RT J. Biol. Chem. 264:6310-6317(1989).
 RN [2]
 RP SEQUENCE OF 1-1154 FROM N.A.
 RA MEDLINE=88112221; PubMed=3338568;
 RT Wood L., Theriault N., Vogell G.;
 RT "cDNA clones completing the nucleotide and derived amino acid
 RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
 RT from mouse.";
 RT FEBS Lett. 227:5-8(1988).
 RN [3]
 RP SEQUENCE OF 1149-1424 FROM N.A.
 RA MEDLINE=86301886; PubMed=3735692;
 RT Nach P., Laurent M., Horn E., Sobel M.E., Zon G., Vogell G.;
 RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
 RT synthetic oligodeoxynucleotide.";
 RT Gene 43:301-304(1986).
 RN [4]
 RP SEQUENCE OF 1276-1669 FROM N.A.
 RA MEDLINE=85127033; PubMed=2578961;
 RT Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
 RT Vogel G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
 RT the alpha 1(IV) chain of basement membrane collagen as derived from
 RT complementary DNA.";
 RT Eur. J. Biochem. 147:217-224(1985).
 RN [5]
 RP SEQUENCE OF 1441-1669 FROM N.A.
 RA MEDLINE=87250460; PubMed=3597383;
 RT Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
 RT Saus J., Pihlajaniemi T.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse
 RT alpha 1(IV) and alpha 2(IV) collagen.";
 RT J. Biol. Chem. 262:8496-8499(1987).
 RN [6]
 RP PARTIAL SEQUENCE FROM N.A.
 RA MEDLINE=86196099; PubMed=3009468;
 RT Sakurai Y., Sullivan M., Yamada Y.;
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
 RT collagen genes.";
 RT J. Biol. Chem. 261:6654-6657(1986).
 RN [7]
 RP SEQUENCE OF 1-28 FROM N.A.
 RA MEDLINE=89066738; PubMed=3198626;
 RT Kaytes P., Wood L., Theriault N., Kurkinen M., Vogell G.;
 RT "Head-to-head arrangement of murine type IV collagen genes.";
 RT J. Biol. Chem. 263:19274-19277(1988).
 RN [8]
 RP SEQUENCE OF 1-28 FROM N.A.
 RA MEDLINE=89071759; PubMed=3200851;
 RT Burbelo P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
 RT bidirectional promoter and a shared enhancer.";
 RT Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).

RN [9]
 RP SEQUENCE OF 1-129 FROM N.A.
 RA MEDLINE=88243724; PubMed=3379041;
 RT Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
 RT collagen chain and the corresponding region of the gene.";
 RT J. Biol. Chem. 263:8706-8709(1988).
 RN [10]
 RP FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC [11]
 CC SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOCFORMS, ALPHA 1(IV) -
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC [12]
 CC DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC [13]
 CC PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC [14]
 CC PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC [15]
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL: J03758; AAA37439.1; -;
 DR EMBL: M23353; AAA51625.1; -;
 DR EMBL: J04694; AAA50292.1; -;
 DR EMBL: X06777; CAA29946.1; -;
 DR EMBL: X02201; CAA26132.1; -;
 DR EMBL: M15832; AAA37340.1; -;
 DR EMBL: M14042; AAA37342.1; -;
 DR EMBL: M12879; AAA37343.1; -;
 DR EMBL: M13024; -; NOT_ANNOTATED_CDS.
 DR EMBL: M13025; -; NOT_ANNOTATED_CDS.
 DR EMBL: M13026; AAA37344.1; -;
 DR EMBL: M13027; AAA37345.1; -;
 DR EMBL: M13043; AAA37346.1; -;
 DR EMBL: J04448; AAA37437.1; -;
 DR PIR: A33525; CGMS4B.
 DR MGI: 88454; Col4a1.
 DR InterPro: IPR001442; C4.
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF01391; Collagen; 21.
 DR Pfam: PF01413; C4; 2.
 DR ProDom: PD003923; C4; 2.
 DR SMART: SM00111; C4; 2.
 DR KX Repeat: Hydroxylation: Glycoprotein; Basement membrane;
 DR KW Repeat: Hydroxylation: Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172
 FT CHAIN 173 1669
 FT DOMAIN 173 1440
 FT DOMAIN 1441 1669
 FT DISULFD 1460 1551
 FT DISULFD 1493 1548
 FT DISULFD 1505 1511
 FT DISULFD 1570 1665
 FT DISULFD 1604 1662
 FT DISULFD 1616 1622
 FT CARBOHYD 126 126
 FT CONFLICT 26 26
 FT CONFLICT 186 186
 FT CONFLICT 319 319
 Amino-terminal propeptide (7S domain).
 Collagen alpha 1(IV) chain.
 Triple-helical region.
 Nonhelical region (NC1).
 OR 1548 (BY SIMILARITY).
 OR 1551 (BY SIMILARITY).
 BY SIMILARITY.
 OR 1662 (BY SIMILARITY).
 OR 1665 (BY SIMILARITY).
 BY SIMILARITY.
 N-LINKED GLYC. . . (POTENTIAL).
 S -> P (IN REF. 2).
 S -> L (IN REF. 2).
 Q -> S (IN REF. 2).

FT CONFLICT 369 369 Q -> L (IN REF. 2).
 FT CONFLICT 403 403 L -> F (IN REF. 2).
 FT CONFLICT 481 481 P -> L (IN REF. 2).
 FT CONFLICT 493 493 Q -> H (IN REF. 2).
 FT CONFLICT 712 712 S -> I (IN REF. 2).
 FT CONFLICT 813 813 E -> Q (IN REF. 2).
 FT CONFLICT 982 982 Q -> H (IN REF. 2).
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).
 SQ SEQUENCE 1669 AA; 160680 MW; 42916891E52058E9 CRC64;
 Query Match 26.8%; Score 49; DB 1; Length 1669;
 Best Local Similarity 39.1%; Pred. No. 1.7e+02;
 Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
 QY 2 GIEGPTLRQMLARRAGPNCIEEP 24
 Db 770 GLTGPPLGIRDPGPGVQGP 792
 1: 11 11 11 11 11
 RESULT 24
 CA17_HUMAN STANDARD; PRT; 2944 AA.
 ID 002388; Q14054; Q16507;
 AC 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE collagen alpha 1(VII) chain precursor (long-chain collagen) (LC COL7A1).
 GN COL7A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94327588; PubMed=8051117;
 RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
 RT "Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain and identification of intragenic polymorphisms."; J. Biol. Chem. 269:20256-20262(1994).
 RN [2]
 RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
 RA MEDLINE=93338437; PubMed=1307247;
 RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G., Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E., Uitto J.;
 RT "The large non-collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor."; Hum. Mol. Genet. 1:475-481(1992).
 RN [3]
 RP SEQUENCE OF 815-1439 FROM N.A.
 RA MEDLINE=9134380; PubMed=1871109;
 RA Parente M.G., Chung L.C., Ryyanen J., Woodley D.T., Wynn K.W., Bauer E.A., Matel M.-G., Chu M.-L., Uitto J.;
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene."; Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
 RN [4]
 RP SEQUENCE OF 369-1255 FROM N.A.
 RA MEDLINE=93107742; PubMed=1469284;
 RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisyah P.S., Cook M.E., Wright J., Briggman R.A., Hunt S.W. III;
 RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion proteins involved in tissue-specific organization of extracellular matrix."; J. Invest. Dermatol. 99:691-696(1992).
 RN [5]
 RP SEQUENCE OF 340-675 FROM N.A.
 RA TISSUE=keratinocytes; PubMed=1567409;
 RA MEDLINE=92231902; PubMed=1567409;
 RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
 RT "Molecular cloning and characterization of type VII collagen cDNA."; Biochem. Biophys. Res. Commun. 183:958-963(1992).
 RN [6]
 RP SEQUENCE OF 2395-2944 FROM N.A.
 RA MEDLINE=93271985; PubMed=8499916;
 RA Greenspan D.S.;
 RT "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene."; Hum. Mol. Genet. 2:273-278(1993).
 RN [7]
 RP SEQUENCE OF 1-87 FROM N.A.
 RA TISSUE=Placenta; PubMed=8088784;
 RA MEDLINE=94375010; PubMed=8088784;
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Uitto J., Greenspan D.S.;
 RT "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene."; Genomics 21:169-179(1994).
 RN [8]
 RP REVIEW ON DEB VARIANTS.
 RA MEDLINE=98041696; PubMed=9375848;
 RA Jaervikallio A., Pulkkinen L., Uitto J.;
 RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen gene (COL7A1)."; Hum. Mutat. 10:338-347(1997).
 RN [9]
 RP VARIANT RDEB LYS-2798.
 RA MEDLINE=93291877; PubMed=8513326;
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;
 RT "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa."; Nat. Genet. 4:62-66(1993).
 RN [10]
 RP VARIANT DDEB SER-2040.
 RA MEDLINE=94224777; PubMed=8170945;
 RA Christiano A.M., Ryyanen M., Uitto J.;
 RT "Dominant dystrophic epidermolysis bullosa: identification of a Gly->Ser substitution in the triple-helical domain of type VII collagen."; Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).
 RN [11]
 RP VARIANT PEB-DDEB CYS-2623.
 RA MEDLINE=96081220; PubMed=8541842;
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;
 RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and identification of a glycine-to-cysteine substitution in the triple-helical domain of type VII collagen."; Hum. Mol. Genet. 4:1579-1583(1995).
 RN [12]
 RP VARIANT DDEB ARG-2043.
 RA MEDLINE=95164985; PubMed=7861014;
 RA Christiano A.M., Morriconi A., Paradisi M., Angelo C., Mazzanti C., Cavallieri R., Uitto J.;
 RT "A glycine-to-arginine substitution in the triple-helical domain of type VII collagen in a family with dominant dystrophic epidermolysis bullosa."; J. Invest. Dermatol. 104:438-440(1995).
 RN [13]
 RP VARIANTS RDEB AND DDEB.
 RA MEDLINE=96220218; PubMed=8644729;
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;
 RT "Glycine substitutions in the triple-helical region of type VII collagen result in a spectrum of dystrophic epidermolysis bullosa phenotypes and patterns of inheritance."; Am. J. Hum. Genet. 58:671-681(1996).
 RN [14]
 RP VARIANT RDEB ARG-2575.
 RA MEDLINE=96154068; PubMed=8592061;
 RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa: genotype/phenotype correlation in a case of moderate clinical

RT severity.";
 RL J. Invest. Dermatol. 106:119-124(1996).
 RN [15]
 RP VARIANT RDEB ARG-1782.
 RX MEDLINE=96183562; PubMed=8618018;
 RA Christiano A.M., McGrath J.A., Uitto J.;
 RT "Influence of the second COL7A1 mutation in determining the
 RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 106:766-770(1996).
 RN [16]
 RP VARIANT RDEB ASP-2073.
 RX MEDLINE=96310789; PubMed=8757758;
 RA Dunhill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
 RA Uitto J., Pope F.M., Eady R.A.J.;
 RT "Clinical and pathological correlations of compound heterozygous COL7A1
 RT mutations in recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 107:171-177(1996).
 RN [17]
 RP VARIANTS RDEB W-1982; G-2008; A-2025; E-2049; G-2063; W-2063 AND
 RP R-2575.
 RX MEDLINE=97465605; PubMed=9326325;
 RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
 RA Freitag S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,
 RA de Prost Y.;
 RT "Characterization of 18 new mutations in COL7A1 in recessive
 RT dystrophic epidermolysis bullosa provides evidence for distinct
 RT molecular mechanisms underlying defective anchoring fibril
 RT formation.";
 RL Am. J. Hum. Genet. 61:599-610(1997).
 RN [18]
 RP VARIANT RDEB ARG-1652.
 RX MEDLINE=98106792; PubMed=9444387;
 RA Cserhalmi-Friedman P.B., Karpali S., Horvath A., Christiano A.M.;
 RT "Identification of a glycine substitution and a splice site mutation
 RT in the type VII collagen gene in a proband with milds recessive
 RT dystrophic epidermolysis bullosa.";
 RL Arch. Dermatol. Res. 289:640-645(1997).
 RN [19]
 RP VARIANTS DEB ARG-2009 AND ARG-2043.
 RX MEDLINE=97358588; PubMed=9215684;
 RA Winberg J.-O., Hammami-Hausali N., Nilsson O., Anton-Lamprecht I.,
 RA Naylor S.L., Kerbacher K., Zimmermann M., Krajci P.,
 RA Gedde-Dahl T., Jr., Bruckner-Tuderman L.;
 RT "Modulation of disease severity of dystrophic epidermolysis bullosa by
 RT a splice site mutation in combination with a missense mutation in the
 RT COL7A1 gene.";
 RL Hum. Mol. Genet. 6:1125-1135(1997).
 RN [20]
 RP VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
 RX MEDLINE=98334662; PubMed=9668111;
 RA Hammami-Hausali N., Schumann H., Raghunath M., Kilgus O., Luethi U.,
 RA Luger T., Bruckner-Tuderman L.;
 RT "Some, but not all, glycine substitution mutations in COL7A1 result in
 RT intracellular accumulation of collagen VII, loss of anchoring
 RT fibrils, and skin blistering.";
 RL J. Biol. Chem. 273:19228-19234(1998).
 RN [21]
 RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.
 RX MEDLINE=98410969; PubMed=9740253;
 RA Kon A., Pulkkinen L., Ishida-Yamamoto A., Hashimoto I., Uitto J.;
 RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis
 RT bullosa.";
 RL J. Invest. Dermatol. 111:534-537(1998).
 RN [22]
 RP VARIANT RDEB ARG-1347.
 RX MEDLINE=99019477; PubMed=9804332;
 RA Terracina M., Posteraro P., Schubert M., Sonego G., Atzori F.,
 RA Zambuno G., Bruckner-Tuderman L., Castiglia D.;
 RT "Compound heterozygosity for a recessive glycine substitution and a
 RT splice site mutation in the COL7A1 gene causes an unusually mild form
 RT of localized recessive dystrophic epidermolysis bullosa.";
 RL J. Invest. Dermatol. 111:744-750(1998).
 RN [23]

RP VARIANTS DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.
 RX MEDLINE=99072663; PubMed=9856843;
 Query Match 26.8%; Score 49; DB 1; Length 2944;
 Best Local Similarity 43.5%; Pred. No. 36+02;
 Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
 QY 2 GIEGPTLRQWLARACPNIGIECP 24
 Db 2218 GLTGPTGANGLPGPSPGLVCP 2240
 RESULT 25
 UROC_MOUSE
 ID UROC_MOUSE STANDARD; PRT; 122 AA.
 AC P81615; 088390.
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Urococtin precursor.
 GN UCN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98292491; PubMed=9628819;
 RA Zhao L., Donaldson C.J., Smith G.W., Vale W.W.;
 RT "The structures of the mouse and human urococtin genes.";
 RL Genomics 50:23-33(1998).
 CC -I- FUNCTION: ACTS IN VITRO TO STIMULATE THE SECRETION OF
 CC ADRENOCORTICOTROPIC HORMONE (ACTH). BINDS WITH HIGH AFFINITY TO
 CC CRF RECEPTOR TYPES 1, 2-ALPHA, AND 2-BETA.
 CC -I- SIMILARITY: BELONGS TO THE SAUVAGINE/CORTICOTROPIN-RELEASING
 CC FACTOR/ROTEMSIN I FAMILY OF PEPTIDES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF038632; AAC24202.1; -
 DR MGD; MGI:1276123; Ucn.
 DR InterPro; IPR00187; CRF.
 DR InterPro; IPR003620; Urococtin_CRF.
 DR Pfam; PF00473; CRF; 1.
 DR ProDom; PD005970; Urococtin_CRF; 1.
 DR SMART; SM00039; CRF; 1.
 DR PROSITE; PS00511; CRF; 1.
 KM Hormone; Amidation; Cleavage on pair of basic residues; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT PROPEP 26 80 BY SIMILARITY.
 FT PEPTIDE 81 120 UROCOCTIN.
 FT MOD_RES 120 120 AMIDATION (G-121 PROVIDE AMIDE GROUP) (BY
 FT SEQUENCE 122 AA; 13557 MW; D2969756F36F5DEA CRC64;
 Query Match 26.5%; Score 48.5; DB 1; Length 122;
 Best Local Similarity 41.7%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 11; Indels 1; Gaps 1;
 QY 6 PIRKWLAAKPGNCIGPTTRQW 29
 Db 21 PESSQWSPAAATGYDPNLR-W 43
 RESULT 26
 CCSL_CAEEL
 ID CCSL_CAEEL STANDARD; PRT; 324 AA.

AC P12114; Q17509;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cuticle collagen scf-1
 GN SQT-1 OR ROL-5 OR B0491.2
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=89028667; PubMed=3180220;
 RA Kramer J.M., Johnson J.J., Edgar R.S., Basch C., Roberts S.;
 RT "The scf-1 gene of C. elegans encodes a collagen critical for
 RT organismal morphogenesis.";
 RL Cell 55:555-565(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RL Sulston J.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
 CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 CC LINKS.
 CC -1- DISASE: THIS IS A COLLAGEN CRITICAL FOR ORGANISMAL MORPHOGENESIS.
 CC MUTATIONS IN SQT-1 CAN LENGTHEN, SHORTEN, OR HELICALLY TWIST THE
 CC ENTIRE ANIMAL.
 CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
 CC COLLAGENS. ROD-6 AND SQT-1 BELONGS TO THE SAME GROUP OF COLLAGEN
 CC AND MAY ALSO PHYSICALLY INTERACT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: J03146; AAA65468.1; -;
 DR EMBL: Z49907; CA900084.1; -;
 DR PIR: A31920; A31920.
 DR WormPep: B0491.2; GE02104.
 DR InterPro: IPR002486; Col_cuticle_N.
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF01391; Collagen: 2.
 DR Pfam: PF01484; Col_cuticle_N: 1.
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
 FT DOMAIN 127 153 TRIPLE-HELICAL REGION.
 FT DOMAIN 171 231 TRIPLE-HELICAL REGION.
 FT DOMAIN 237 299 TRIPLE-HELICAL REGION.
 FT CONFLICT 158 158 V -> A (IN REF. 2).
 FT CONFLICT 238 238 G -> R (IN REF. 2).
 FT SEQUENCE 324 AA; 32779 MW; DBAC0082699301CFC CRC64;
 Query Match 26.5%; Score 48.5; DB 1; Length 324;
 Best Local Similarity 45.8%; Pred. No. 39;
 Matches 11; Conservative 4; Mismatches 6; Indels 3; Gaps 1;
 QY 2 GIEGPTLRQWLARAGPNGIEGPT 25
 DB 270 GPEPSSGKQ---GRGGPDGTGGFT 290
 RESULT 27
 SYK_AERPE STANDARD; PRT; 562 AA.
 ID SYK_AERPE
 AC G9YFT9;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lysyl-tRNA synthetase (PC 6.1.1.6) (Lysine--tRNA ligase) (LYSRS).
 GN LYS OR APE0161.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Karabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankei A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kudota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 CC -1- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) -> AMP + diphosphate
 CC + L-lysyl-tRNA(Lys).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AP000058; BAA79072.1; -;
 DR InterPro: IPR001412; tRNA-synt_1.
 DR InterPro: IPR002904; tRNA-synt_1f_1.
 DR Pfam: PF01921; tRNA-synt_1f: 1.
 DR PROSITE: PS00178; AA-tRNA-LIGASE.1; FALSE_NEG.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 50 58 "HIGH" REGION.
 FT SITE 305 309 "KMSKS" REGION.
 FT SEQUENCE 562 AA; 65114 MW; 753664E2937FBE27 CRC64;
 Query Match 26.5%; Score 48.5; DB 1; Length 562;
 Best Local Similarity 34.2%; Pred. No. 67;
 Matches 13; Conservative 5; Mismatches 11; Indels 9; Gaps 2;
 QY 2 GIEGP--TLRQWLARAG-----PNGIEGPTLRQWL 30
 DB 283 GFEPGWEYEWVSLRAGRGREADWSSGFGTTPREWL 320
 RESULT 28
 Vxis_BP434 STANDARD; PRT; 72 AA.
 ID Vxis_BP434
 AC P11683; P16408;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Excisionase.
 GN XIS.
 OS Bacteriophage 434, and
 OS Bacteriophage HK022.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
 OC Lambda phage group
 OX NCBI_TaxID=10712, 10742;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=phage 434;
 RX MEDLINE=91346141; PubMed=1715186;

Wed Oct 9 10:29:43 2002

us-09-422-838c-25.rsp

Page 20

Sequence 270 AA; 30984 MW; 1EC5ACDEDD8AB92 CRC64;

Query Match 26.2%; Score 48; DB 1; Length 270;

Best Local Similarity 31.4%; Pred. No. 38;

Matches 11; Conservative 5; Mismatches 9; Indels 10; Gaps 2;

QY 7 TLROMLAARAGP-----NGIEGPT-LROMLA 31

DB 151 TLQAMLVGHKGPLAKKPKQHLSVDNPTITGRWLA 185

Search completed: October 9, 2002, 09:00:12
Job time : 6.14422 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:52:16 ; Search time 12.1827 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-25

Perfect score: 183
Sequence: 1 GGIEGPTLRQMLAARAGNGIEGPTLRQMLARA 34

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_19:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rodent:*
13: sp.virus:*
14: sp.vertibrate:*
15: sp.unclassified:*
16: sp.rvrius:*
17: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	34.4	683	16	083436 treponema p
2	62	33.9	607	2	Q918D4
3	60	32.8	509	2	Q9S5E5
4	58.5	32.0	869	5	Q9VZ82
5	58	31.7	214	5	Q20968
6	56	30.6	337	10	Q24514
7	56	30.6	361	16	Q9ABCT
8	55.5	30.3	246	13	Q919S8
9	55.5	30.3	246	13	Q919S7
10	55.5	30.3	1744	3	Q94192
11	55	30.1	420	2	P97011
12	55	30.1	902	5	Q16161
13	54	29.5	250	10	Q9AS26
14	54	29.5	1095	16	Q91304
15	54	29.5	1366	4	Q15177
16	54	29.5	3198	5	Q26639

17	53.5	29.2	246	13	Q91872
18	53.5	29.2	246	13	Q919T0
19	53.5	29.2	246	13	Q919S6
20	53.5	29.2	371	16	Q91477
21	53	29.0	305	2	Q9S0M9
22	53	29.0	326	16	Q9RTE6
23	53	29.0	967	2	Q9K2D5
24	53	29.0	1433	11	Q07563
25	53	29.0	1820	13	Q91907
26	52.5	28.7	333	10	Q94LX0
27	52.5	28.7	539	2	Q9RK76
28	52.5	28.7	814	4	Q96C78
29	52	28.4	214	5	Q20964
30	52	28.4	230	5	Q61518
31	52	28.4	395	2	Q9X7N5
32	52	28.4	1349	2	Q91096
33	51.5	28.1	243	13	Q91902
34	51.5	28.1	390	13	Q91903
35	51	27.9	215	5	Q20967
36	51	27.9	264	2	Q93J78
37	51	27.9	281	17	Q9YD00
38	51	27.9	289	2	Q93P85
39	51	27.9	306	16	Q05576
40	51	27.9	322	2	Q9RK51
41	51	27.9	326	2	P95613
42	51	27.9	366	10	Q943T6
43	51	27.9	381	2	Q9X757
44	51	27.9	589	5	Q18756
45	51	27.9	591	4	Q96HC0

ALIGNMENTS

RESULT 1
ID 083436 PRELIMINARY; PRT: 683 AA.
AC 083436;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0421.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ullrich T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete."
RL Science 281:375-388(1998).
DR EMBL: AE001220; AAC65409.1; .
DR TIGR: TP0421; .
DR InterPro: IPR001258; NHL.
DR InterPro: IPR01440; TPR.
DR Pfam: PF01436; NHL; 4.
DR Pfam: PF00515; TPR; 1.
KW Complete proteome.
SQ SEQUENCE 683 AA; 74518 MW; F91407FA7094AADI CRC64;
Query Match 34.4%; Score 63; DB 16; Length 683;
Best Local Similarity 46.4%; Pred. NO. 6.5;
Matches 13; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

```

QY 6 PTLRQMLAARAGPNCIGPTRLQMLAAR 33
   1 : ||| ||| ||| |||
DB 74 PLELEWGNMYRSGICGALHQMGAAR 101

RESULT 2
ID 09LBD4 PRELIMINARY; PRT; 607 AA.
AC 09LBD4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 66.3 KDA PROTEIN (FRAGMENT).
OS Polyangium cellulolum.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SO CE90;
RX MEDLINE=20130945; PubMed=10662695;
RA Molnar I., Schupp T., Ono M., Zirkle R.E., Milnamow M.,
RA Nowak-Thompson B., Engel N., Toupet C., Strattmann A., Cyr D.D.,
RA Grolach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RA "The biosynthetic gene cluster for the microtubule-stabilizing agents
RA epothilones A and B from Sorangium cellulosum So ce90."
RL Chem. Biol. 7:97-109(2000).
DR EMBL: AF210843; AAF26904.1; -
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 607 AA; 66326 MW; F113CA299825048E CRC64;

Query Match
Best Local Similarity 33.9%; Score 62; DB 2; Length 607;
Matches 16; Conservative 3; Mismatches 7; Indels 20; Gaps 2;

QY 3 IEPTLRQMLAARAGPNCIGP-----TLRQMLAA 32
   1 : ||| ||| ||| |||
DB 96 VDGPAVLVRLAARAGP-----GLREYEEERERARTAOEARRLWAA 137

RESULT 3
ID 09S5E5 PRELIMINARY; PRT; 509 AA.
AC 09S5E5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE ORF1, ORF2, ORF3, ORF4, ORF5 GENES, COMPLETE CDS (PUTATIVE DNA-BINDING
DE PROTEIN).
GN SC9H11.15.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA Umeyama T., Ping Chin L., Horinouchi S.;
RT "Multicopy suppressor gene of afsR mutant."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);

```

```

RX MEDLINE=97000351; PubMed=883436;
RA Kienbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AB017438; BAA82701.1; -
DR EMBL: AL356592; CAB92204.1; -
KW DNA-binding.
SQ SEQUENCE 509 AA; 54398 MW; 7BB074DAAE0F1867 CRC64;

Query Match
Best Local Similarity 44.1%; Score 60; DB 2; Length 509;
Matches 15; Conservative 4; Mismatches 11; Indels 4; Gaps 2;

QY 3 IEPTLRQMLAARAGPNCIGPTRLQMLAAR 32
   1 : ||| ||| ||| |||
DB 404 LAGPALRTWAVDLGRDPDGRDLRRLRTLTWIAA 437

RESULT 4
ID 09VZ82 PRELIMINARY; PRT; 869 AA.
AC 09VZ82;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CG7479. PROTEIN.
GN CG7479.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisli M., Kaul S.F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska Z., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

RA Gldbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003482; AAF47943.1; -
 DR FlyBase: FBgn0035576; CG7479.
 DR InterPro: IPR002300; tRNA-synt_1a.
 DR InterPro: IPR004412; tRNA-synt_1.
 DR InterPro: IPR002302; tRNA-synt_1.
 DR Pfam: PF00133; tRNA-synt_1.1.
 DR PRINTS: PR00985; TRNASYNTHLEU.
 DR PROSITE: PS00178; AA.TRNA.LIGASE.I.1.
 SQ SEQUENCE 869 AA; 99299 MW; E87A1ECBB27867 CRC64;

Query Match 32.0%; Score 58.5; DB 5; Length 869;
 Best Local Similarity 40.6%; Pred. No. 31;
 Matches 13; Conservative 4; Mismatches 10; Indels 5; Gaps 1;

QY 3 IEGLPTLRQWLA-----ARAGPNGIEGPTLRQW 29
 DB 213 VEKRLRQWFIKTSYAKKOLLIDLEPTLRDW 244

RESULT 5
 ID Q20968 PRELIMINARY; PRT; 214 AA.
 AC Q20968;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE F58B3.3 PROTEIN.
 GN F58B3.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RE SEQUENCE FROM N.A.
 RA Harris B.R.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C.elegans*: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z73427; CAA97801.1; -
 SQ SEQUENCE 214 AA; 23089 MW; F41992EC471FF165 CRC64;

Query Match 31.7%; Score 58; DB 5; Length 214;
 Best Local Similarity 47.6%; Pred. No. 8.1;
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 GGIETPLRQWLAARAGPNCI 21
 DB 186 GGWSPTHQWEGTGAAGPCGV 206

RESULT 6
 ID Q24514 PRELIMINARY; PRT; 337 AA.
 AC Q24514;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ANI1.
 GN ANI1.
 OS *Petunia hybrida* (petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
 OX NCBI_TaxID=4102;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=CV V26; TISSUE=COROLLA;
 RX MEDLINE=97336075; PubMed=9192870;
 RA de Vetten N., Quattrocchio F., Mol J., Koes R.;
 RT "The anil locus controlling flower pigmentation in petunia encodes a
 RT novel WD-repeat protein conserved in yeast, plants, and animals."
 RL Genes Dev. 11:1422-1434(1997).
 CC -1 SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: U94748; AAC18914.1; -
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 4.
 DR PRINTS: PR00320; GPROTEINBRPT.
 DR SMART: SM00320; WD40; 3.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_2.
 DR PROSITE: PS00678; WD_REPEATS_1; 2.
 DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 337 AA; 37857 MW; 7024CAED5FD109C CRC64;

Query Match 30.6%; Score 56; DB 10; Length 337;
 Best Local Similarity 31.9%; Pred. No. 23;
 Matches 15; Conservative 3; Mismatches 13; Indels 16; Gaps 1;

QY 1 GGIETPLRQWLAARAGPNCIE-----GPTLRQWLA 31
 DB 279 GGDDGQALIMELPTVAGPNDPMWSYAGAEINQLQNSPARDWIA 325

RESULT 7
 ID Q9ABC7 PRELIMINARY; PRT; 361 AA.
 AC Q9ABC7;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE CATTON EFFLUX FAMILY PROTEIN.
 GN CC0303.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RE SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nieman W.C., Feldblum T.V., Laud M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Uitterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AE005704; AAK22290.1; -
 DR TIGR: CC0303; -
 DR InterPro: IPR002524; Cation_efflux.
 DR InterPro: IPR002395; Kininogen.
 DR Pfam: PF01545; Cation_efflux_1.
 DR PRINTS: PR00334; KININOGEN.
 KW Complete proteome.
 SQ SEQUENCE 361 AA; 38180 MW; 1A4F7FOA7C62EEB0 CRC64;

Query Match 30.6%; Score 56; DB 16; Length 361;
 Best Local Similarity 54.5%; Pred. No. 25;
 Matches 12; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 12 LAARAGPNCIEGPTLRQWLAAR 33
 DB 266 LALDAPRGIDTQKVRDMLAR 287

RESULT 8

0919S8
ID 0919S8 PRELIMINARY; PRT; 246 AA.
AC 0919S8;
DT 01-OCT-2000 (TREMBLREL. 15, Created)
DT 01-OCT-2000 (TREMBLREL. 15, last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, last annotation update)
DE GAG POLYPROTEIN (FRAGMENT).
CN GAG.
OS Phasianus colchicus (Ring-necked pheasant).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Phasianus.
OX NCBI_TaxID=9054;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20219390; PubMed=10756010;
RA Dimcheff D.E., Drovetski S.V., Krishnan M., Mindell D.P.;
RT "Cospeciation and horizontal transmission of avian sarcoma and
leukosis virus gag genes in galliform birds.";
RL J. Virol. 74:3984-3995(2000).
DR EMBL: AF225386; AAF64756.1; -.
DR HSSP: P03322; 1A6S.
DR InterPro: IPR004028; Retro.M.
DR Pfam: PF02813; Retro.M; 1.
KM Polyprotein.
FT NON_TER 1 1
SQ SEQUENCE 246 AA; 24950 MW; 5EA4B3247359A987 CRC64;

Query Match 30.3%; Score 55.5; DB 13; Length 246;
Best Local Similarity 34.8%; Pred. No. 19;
Matches 16; Conservative 2; Mismatches 15; Indels 13; Gaps 2;

QY 1 GGIGPILROWLAARAG-PNGIE-----GPTLRQWLAR 33
DB 178 GSPGEGVATSLAGRGPPRGVEQRAEPGCPDAPGALTDMWRIR 223

RESULT 9
0919S7
ID 0919S7 PRELIMINARY; PRT; 246 AA.
AC 0919S7;
DT 01-OCT-2000 (TREMBLREL. 15, Created)
DT 01-OCT-2000 (TREMBLREL. 15, last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, last annotation update)
DE GAG POLYPROTEIN (FRAGMENT).
CN GAG.
OS Phasianus colchicus (Ring-necked pheasant).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Phasianus.
OX NCBI_TaxID=9054;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20219390; PubMed=10756010;
RA Dimcheff D.E., Drovetski S.V., Krishnan M., Mindell D.P.;
RT "Cospeciation and horizontal transmission of avian sarcoma and
leukosis virus gag genes in galliform birds.";
RL J. Virol. 74:3984-3995(2000).
DR EMBL: AF225387; AAF64757.1; -.
DR HSSP: P03322; 1A6S.
DR InterPro: IPR004028; Retro.M.
DR Pfam: PF02813; Retro.M; 1.
KM Polyprotein.
FT NON_TER 1 1
SQ SEQUENCE 246 AA; 24977 MW; DF208DF53E58B789 CRC64;

Query Match 30.3%; Score 55.5; DB 13; Length 246;
Best Local Similarity 34.8%; Pred. No. 19;
Matches 16; Conservative 3; Mismatches 14; Indels 13; Gaps 2;
QY 1 GGIGPILROWLAARAG-PNGIE-----GPTLRQWLAR 33

DB 178 GSPGEGVATSLAGRGPPRGVEQRAEPGCPDAPGALTDMWRIR 223

RESULT 10
094192
ID 094192 PRELIMINARY; PRT; 1744 AA.
AC 094192;
DT 01-MAY-1999 (TREMBLREL. 10, Created)
DT 01-JUN-2001 (TREMBLREL. 17, last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, last annotation update)
DE CHITIN SYNTHASE.
CN CHS4.
OS Paracoccidioides brasiliensis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Onygenales; mitosporic Onygenales; Paracoccidioides.
OX NCBI_TaxID=121759;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20210320; PubMed=10746225;
RA Nino-Vega G.A., Munro C.A., San-Blas G., Gooday G.W., Gow N.A.;
RT "Differential expression of chitin synthase genes during temperature-
induced dimorphic transitions in Paracoccidioides brasiliensis.";
RL Med. Mycol. 38:31-39(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Nino-Vega G.A., San-Blas G.;
RT "Sequence analysis of the CHS4 gene of Paracoccidioides
brasiliensis.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF107624; AAD19613.2; -.
DR InterPro: IPR002923; Chitin_synth.
DR InterPro: IPR001177; Cu-oxidase.
DR InterPro: IPR001173; Glycos.transf-2.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF03142; Chitin_synth_2; 1.
DR Pfam: PF00063; myosin_head; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
SQ SEQUENCE 1744 AA; 193777 MW; DB7622D0A69F0705 CRC64;

Query Match 30.3%; Score 55.5; DB 3; Length 1744;
Best Local Similarity 51.7%; Pred. No. 1.5e+02;
Matches 15; Conservative 2; Mismatches 11; Indels 1; Gaps 1;
QY 7 TLROWL-AARAGPNGIEGPTLRQWLARA 34
DB 56 TWTWLTAAAPSPNGEVGCTIDADLARRA 84

RESULT 11
P97011
ID P97011 PRELIMINARY; PRT; 420 AA.
AC P97011;
DT 01-MAY-1997 (TREMBLREL. 03, Created)
DT 01-MAY-1997 (TREMBLREL. 03, last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, last annotation update)
DE SORBITOL OXIDASE.
CN SOX.
OS Streptomyces sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyicinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1931;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H-7775;
RA Hiraga K., Eto T., Yoshioka I., Oda K.;
RT "Cloning of a gene encoding a sorbitol oxidase from Streptomyces sp.
H-7775.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB000519; BAA19135.1; -.
DR InterPro: IPR001575; Oxid_FAD_bind.
DR Pfam: PF01565; FAD_binding_4; 1.

SO SEQUENCE 420 AA; 45181 MW; EF3189045CA0649 CRC64;
Query Match 30.1%; Score 55; DB 2; Length 420;
Best Local Similarity 37.9%; Pred. No. 39;
Matches 11; Conservative 2; Mismatches 16; Indels 0; Gaps 0;
QY 5 GPTLRQWLARAGPNGIEGPTLRQWLAR 33
DB 215 GPVGQWLKRVGDEGASVMPAEMLGAR 243
RESULT 12
ID 016161 PRELIMINARY; PRT; 902 AA.
AC 016161;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRECOLLAGEN P PRECURSOR.
GN PRECOL-P.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FOOT;
RX MEDLINE=97442537; PubMed=9295275;
RA Coyne K.J., Qin X.X., Waite J.H.;
RT "Extensile collagen in mussel byssus: A natural block copolymer";
DR EMBL; AF015539; AAB80719.1; -;
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 7.
KW Signal; Collagen.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 902 COLLAGEN P.
FT SEQUENCE 902 AA; 78526 MW; D1E09DEABD9EF3 CRC64;
Query Match 30.1%; Score 55; DB 5; Length 902;
Best Local Similarity 52.2%; Pred. No. 89;
Matches 12; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 2 GIEGPTLRQWLARAGPNGIEGP 24
DB 525 GPKGPTGAGPAGPAGPSGEGCP 547
RESULT 13
ID 09AS26 PRELIMINARY; PRT; 250 AA.
AC 09AS26;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE P0416G11.19 PROTEIN.
GN P0416G11.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0416G11.1";
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002968; BAB39267.1; -;
DR SQUENCE 250 AA; 26094 MW; 617AB53A7738C4B7 CRC64;
Query Match 29.5%; Score 54; DB 10; Length 250;

Best Local Similarity 35.6%; Pred. No. 30;
Matches 16; Conservative 2; Mismatches 15; Indels 12; Gaps 2;
QY 2 GIEGPTLRQWLARAGPNGIEGPTLRQWLAR 34
DB 43 GAGPGPKEWPGCSLTVHGPGAPGLTPAGAVGPTROLHLARA 87
RESULT 14
ID 091304 PRELIMINARY; PRT; 1095 AA.
AC 091304;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROBABLE PYRUVATE CARBOXYLASE.
GN PA1400.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen";
RT Nature 406:959-964(2000).
CC -1. CORFACTOR: BIOTIN (BY SIMILARITY).
DR EMBL; AE004569; AAG04789.1; -;
DR HSSP; P24182; 1BNC.
DR InterPro: IPR001249; AccoA_biotinC.
DR InterPro: IPR001882; Biotin.
DR InterPro: IPR000089; Biotin_lipoyl.
DR InterPro: IPR000022; Carboxyl_trans.
DR InterPro: IPR000901; CPSase.
DR InterPro: IPR001064; crystallin.
DR Pfam: PF02785; Biotin_carb.C; 1.
DR Pfam: PF00364; Biotin_lipoyl; 1.
DR Pfam: PF01039; Carboxyl_trans; 1.
DR Pfam: PF00289; CPSase_L_chain; 1.
DR Pfam: PF02786; CPSase_L_D2; 1.
DR PROSITE; PS00186; BIOTIN; 1.
DR PROSITE; PS00867; CPSASE_2; UNKNOWN_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Biotin; Complete proteome; Pyruvate.
SO SEQUENCE 1095 AA; 116876 MW; 34370FB8BEC201AD CRC64;
Query Match 29.5%; Score 54; DB 16; Length 1095;
Best Local Similarity 45.5%; Pred. No. 1.5e+02;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
QY 3 IEGPTLRQWLARAGPNGIEGP 24
DB 786 IEGGLGRFAAEVGPVGQGP 807
RESULT 15
ID 015177 PRELIMINARY; PRT; 1366 AA.
AC 015177;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PREPRO-ALPHA2(I) COLLAGEN PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ID 0919T0 PRELIMINARY; PRT; 246 AA.
 AC 0919T0;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE GAG POLYPROTEIN (FRAGMENT).
 GN GAG.
 OS Phasianus colchicus (Ring-necked pheasant).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Phasianus.
 OX NCBI_TaxID=9054;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20219390; PubMed=10756010;
 RA Dimcheff D.E., Drovetski S.V., Krishnan M., Mindell D.P.;
 RT "Cospeciation and horizontal transmission of avian sarcoma and
 RT leukosis virus gag genes in galliform birds."
 RL J. Virol. 74:3984-3995(2000).
 DR EMBL; AF225383; AAF64753.1;-.
 DR HSSP; P03322; 1A6S.
 DR InterPro; IPR004028; Retro_M.
 DR Pfam; PF02813; Retro_M.1.
 KW Polyprotein.
 FT NON_TER 1 1
 FT 246 246
 SQ SEQUENCE 246 AA; 25036 MW; 2F8B52D95FBD316 CRC64;

Query Match 29.2%; Score 53.5; DB 13; Length 246;
 Best Local Similarity 34.8%; Pred. No. 34;
 Matches 16; Conservative 2; Mismatches 15; Indels 13; Gaps 2;

OY 1 GGIEGPTLRQWLARAG-PNGIE-----GPTLRQWLAR 33
 DB 178 GGPEGEVATSLAGRDGPPRGVQRARPCGCPDAPGALDMARIR 223

RESULT 19
 0919S6 PRELIMINARY; PRT; 246 AA.
 AC 0919S6;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE GAG POLYPROTEIN (FRAGMENT).
 GN GAG.
 OS Phasianus colchicus (Ring-necked pheasant).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OX NCBI_TaxID=9054;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20219390; PubMed=10756010;
 RA Dimcheff D.E., Drovetski S.V., Krishnan M., Mindell D.P.;
 RT "Cospeciation and horizontal transmission of avian sarcoma and
 RT leukosis virus gag genes in galliform birds."
 RL J. Virol. 74:3984-3995(2000).
 DR EMBL; AF225383; AAF64753.1;-.
 DR HSSP; P03322; 1A6S.
 DR InterPro; IPR004028; Retro_M.
 DR Pfam; PF02813; Retro_M.1.
 KW Polyprotein.
 FT NON_TER 1 1
 FT 246 246
 SQ SEQUENCE 246 AA; 25028 MW; 25723F710FE374A CRC64;

Query Match 29.2%; Score 53.5; DB 13; Length 246;
 Best Local Similarity 34.8%; Pred. No. 34;
 Matches 16; Conservative 2; Mismatches 15; Indels 13; Gaps 2;
 OY 1 GGIEGPTLRQWLARAG-PNGIE-----GPTLRQWLAR 33
 DB 178 GGPEGEVATSLAGRDGPPRGVQRARPCGCPDAPGALDMARIR 223

DB 178 GGPEGEVATSLAGRDGPPRGVQRARPCGCPDAPGALDMARIR 223

RESULT 20
 091477 PRELIMINARY; PRT; 371 AA.
 AC 091477;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN PA1267.
 GN PA1267.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Huinagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolenino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.P.,
 RA Reizer J., Saiter M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RT Nature 406:959-964(2000).
 RL EMBL; AE004556; AAG04656.1;-.
 DR InterPro; IPR00205; NAD_binding.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 371 AA; 39174 MW; 016D60440AD50D7 CRC64;

Query Match 29.2%; Score 53.5; DB 16; Length 371;
 Best Local Similarity 29.8%; Pred. No. 53;
 Matches 14; Conservative 8; Mismatches 12; Indels 13; Gaps 2;

OY 1 GGIEGPTLRQWLARAG-PNGIEPTLR-----QWLARA 34
 DB 140 GILYAPMARWLDQAGPRLRLYAFVSEVDSRLADGRWLASAE 186

RESULT 21
 09S0M9 PRELIMINARY; PRT; 305 AA.
 AC 09S0M9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE UV-ENDORNUCLEASE.
 GN UVSCDE.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KRL;
 RA Kitayama S., Kikuchi M., Funayama T., Narumi I., Watanabe H.;
 RT "Cloning of structural gene of an alternative incision enzyme for DNA
 RT damage in Deinococcus radiodurans."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB033747; BAA85759.1;-.
 KW Endonuclease.
 SQ SEQUENCE 305 AA; 33592 MW; B94D333243E2FE4 CRC64;

Query Match 29.0%; Score 53; DB 2; Length 305;
 Best Local Similarity 40.5%; Pred. No. 50;
 Matches 15; Conservative 4; Mismatches 6; Indels 12; Gaps 2;
 OY 4 EGPTLRQWLARAG-PNGIEGPTLRQ 28
 DB 228 EDPSTVREWLARARATWPPPMQVYHLSNGIEGPODR 264

```

RESULT 22
Q9RTE6 PRELIMINARY; PRT: 326 AA.
ID Q9RTE6
AC Q9RTE6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE UV DAMAGE ENDONUCLEASE, PUTATIVE.
GN DR1819.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=2003696; PubMed=10567266;
RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Moffat R.U., Halt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Dodgson K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamatheva R.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarov A.K.S., Asavind L., Daly M.J., Martin K.W., Feilichmann R.D.,
RA Keithum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).
DR EMBL, AE002022; NAF11370.1; -.
DR TIGR, DR1819; -.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 326 AA; 35693 MW; CAEAD0DADZC38988 CRC64;

Query Match 29.0%; Score 53; DB 16; Length 326;
Best Local Similarity 40.5%; Pred. No. 53;
Matches 15; Conservative 4; Mismatches 6; Indels 12; Gaps 2

QY 4 EGGTLROW-LAARG-----PNCIEGPTLRQ 28
|::||:|||| |::||:|
Db 249 EDPSVRENYLARATWOPPEMWYVHNSNGEGSPQDR 285

RESULT 23
Q9KZD5 PRELIMINARY; PRT: 967 AA.
ID Q9KZD5
AC Q9KZD5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE PROBABLE NADH DEHYDROGENASE I COMPLEX, SUBUNIT.
GN SC6F7.07.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Regenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RL "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
```



```

DE PRO-ALPHA 1 TYPE V/XI COLLAGEN.
GN COLV/XI.1.
OS Pagrus major (Red sea bream) (Chrysophrys major).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Sparidae; Pagrus
OX NCBI_TaxID=143350;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21240220; PubMed=11342118;
RA Tsubata K., Tanaka H., Yokoyama Y., Sakaguchi M., Toyohara H.;
RT "Structure of a full-length cDNA clone for the pro-1(V/XI) collagen
RL chain of red seabream."
RL Biochim. Biophys. Acta 1517:323-326(2001).
DR EMBL; AB045975; BAB03287.1; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001230; Prenyltn.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD002078; Fib.collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW Collagen.
SQ SEQUENCE 1820 AA; 181678 MW; 46E45E8AF7AD3DAE CRC64;

Query Match 29.0%; Score 53; DB 13; Length 1820;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 GIEGPTLRQWLAARAGPENGIEG 23
| | | | | | | | | | | | | |
Db 1400 GKTGPVGPQGLAKAGEGEGRG 1421

RESULT 26
Q94LX0 PRELIMINARY; PRT; 333 AA.
AC Q94LX0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE PUTATIVE REGULATORY PROTEIN IN ANTHOCYANIN BIOSYNTHESIS.
OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.
OX NCBI_TaxID=48386;
RN [1]
RP SEQUENCE FROM N.A.
RA Somporipailin K., Makita Y., Yamazaki M., Saito K.;
RT "A two-repeat-containing putative regulatory protein in anthocyanin
RL biosynthesis in Perilla frutescens."
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AA059642; BAB58883.1; -
SQ SEQUENCE 333 AA; 36844 MW; A52E156D6843C8AA CRC64;

Query Match 28.7%; Score 52.5; DB 10; Length 333;
Best Local Similarity 36.6%; Pred. No. 63;
Matches 15; Conservative 5; Mismatches 12; Indels 9; Gaps 1;

QY 2 GIEGPTLRQWLAARAGPENGIEGPTL-----RQWLAAR 33
| : : : | | | | | | | | | | | | | |
Db 276 GDDGOSLMLPTVAGPNCIDPMYMSAGAEINQLOWNSAAQ 316

RESULT 27
Q9RK76

```

```

ID Q9RK76 PRELIMINARY; PRT; 539 AA.
AC Q9RK76;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE BETA-HEXOSAMINIDASE.
GN SCF11.14.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K.J., Harris D.;
RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RC MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalre D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL132662; CAB59591.1; -
DR HSSP; P06865; IQBC.
DR InterPro; IPR001540; Glyco_hydro_20.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00728; Glyco_hydro_20; 1.
DR Pfam; PF02858; Glyco_hydro_20b; 1.
DR PRINTS; PR00738; GLHYDRLASE20.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 539 AA; 58624 MW; 0861929C641DAC56 CRC64;

Query Match 28.7%; Score 52.5; DB 2; Length 539;
Best Local Similarity 41.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 2; Mismatches 14; Indels 7; Gaps 3;

QY 1 GGIEGPTLRQW-----LAARAGPENGIEG-TRQWLAAR 33
| | | | | | | | | | | | | |
Db 311 GGDEVT-TEWELSPARARARREGIAGRALHPWFIAR 348

RESULT 28
Q96C78 PRELIMINARY; PRT; 814 AA.
AC Q96C78;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 15 (METARCIDIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=KIDNEY, AND RENAL CELL ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014566; AAH14566.1; -
KW Integrin.
SQ SEQUENCE 814 AA; 87717 MW; 683A8368AD3096B CRC64;

Query Match 28.7%; Score 52.5; DB 4; Length 814;
Best Local Similarity 44.8%; Pred. No. 1.6e+02;
Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

```

OY 3 IEGPTLRQWLARAGPNGIEPTLRQWLA 31
 Db 728 LKGPFC-QYRAOSGSPSERGPGPRALMLA 755

RESULT 29

ID 020964 PRELIMINARY: PRT: 214 AA.
 AC 020964;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE F58B3.1 PROTEIN.
 GN F58B3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.R.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: Z73427; CA97797.1; -;
 DR EMBL: Z73427; CA97797.1; -;
 SQ SEQUENCE 214 AA; 23608 MM; 63C30FDE325049F3 CRC64;

Query Match 28.4%; Score 52; DB 5; Length 214;
 Best Local Similarity 42.9%; Pred. No. 46;
 Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

OY 1 GGIEPTLRQWLARAGPNGI 21

Db 186 GGSRRPTTHQWEGTTGPGCV 206

RESULT 30

ID 061518 PRELIMINARY: PRT: 230 AA.
 AC 061518;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL 25.8 KDA PROTEIN.
 GN P17E9.11.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX Woessner J.;
 RT "The sequence of C. elegans cosmid P17E9.";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;

RT "Direct Submission";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF047656; AAC05110.1;
 DR InterPro: IPR003006; Iq_MHC.
 DR PROSITE: PS00290; Iq_MHC; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 230 AA; 25772 MM; 48DB7079872B9432 CRC64;

Query Match 28.4%; Score 52; DB 5; Length 230;
 Best Local Similarity 42.9%; Pred. No. 49;
 Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 GGIEPTLRQWLARAGPNGI 21
 Db 202 GGMKPTTHQWEGTTGPGCV 222

Search completed: October 9, 2002, 09:03:07
 Job time : 13.266 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 ; Search time 16.1874 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-26

Perfect score: 194

Sequence: 1 IECPTRLRWLAARAGGGGGGIEGPTLRWLARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	194	100.0	36	21	AA16963
2	194	100.0	36	21	AA17293
3	194	100.0	36	21	AA196525
4	194	100.0	41	21	AA196528
5	194	100.0	42	21	AA17281
6	194	100.0	42	21	AA17282
7	194	100.0	42	21	AA17308
8	194	100.0	42	21	AA196530
9	194	100.0	60	21	AA17311
10	194	100.0	269	21	AA16960
11	194	100.0	269	21	AA196531

12	190	97.9	268	21	AA16959	Fc-TMP-TMP protein
13	186	95.9	36	21	AA17301	TPO-mimetic peptide
14	186	95.9	36	21	AA196523	Thrombopoietin mim
15	185	95.4	36	21	AA17303	TPO-mimetic peptide
16	185	95.4	36	21	AA17307	TPO-mimetic peptide
17	185	95.4	36	21	AA196524	Thrombopoietin mim
18	183.5	94.6	37	21	AA17294	TPO-mimetic peptide
19	183	94.3	38	21	AA17295	TPO-mimetic peptide
20	182.5	94.1	39	21	AA17304	TPO-mimetic peptide
21	182.5	94.1	39	21	AA17305	TPO-mimetic peptide
22	182	93.8	36	21	AA17306	TPO-mimetic peptide
23	182	93.8	36	21	AA196526	Thrombopoietin mim
24	181	93.3	42	21	AA17296	TPO-mimetic peptide
25	177.5	91.5	35	21	AA17297	TPO-mimetic peptide
26	174	89.7	40	21	AA17302	TPO-mimetic peptide
27	171	88.1	34	21	AA17291	TPO-mimetic peptide
28	168	86.6	36	21	AA17298	TPO-mimetic peptide
29	168	86.6	36	21	AA17299	TPO-mimetic peptide
30	168	86.6	36	21	AA196521	Cyclic or linear t
31	166	85.6	36	21	AA17300	TPO-mimetic peptide
32	166	85.6	36	21	AA196522	Linear thrombopo
33	164.5	84.8	33	21	AA17290	TPO-mimetic peptide
34	158	81.4	32	21	AA17289	TPO-mimetic peptide
35	151.5	78.1	31	21	AA17288	TPO-mimetic peptide
36	145	74.7	30	21	AA17287	TPO-mimetic peptide
37	144	74.2	32	21	AA17297	TPO-mimetic peptide
38	144	74.2	32	21	AA196520	Thrombopoietin mim
39	144	74.2	34	21	AA196527	Thrombopoietin mim
40	138.5	71.4	29	21	AA17286	TPO-mimetic peptide
41	132	68.0	28	21	AA17285	TPO-mimetic peptide
42	131.5	67.8	29	21	AA16970	TPO-mimetic peptide
43	129.5	66.8	31	21	AA16973	TPO-mimetic peptide
44	129.5	66.8	31	21	AA16974	TPO-mimetic peptide
45	125.5	64.7	29	21	AA16971	TPO-mimetic peptide

ALIGNMENTS

RESULT 1

AA16963	1	AA16963 standard; Protein; 36 AA.
ID	AA16963	
XX	AA16963	
AC	AA16963	
XX	AA16963	
DT	31-OCT-2000 (first entry)	
XX	31-OCT-2000	
XX	31-OCT-2000	
DE	TPO-mimetic peptide TMP-TMP SEQ ID NO:14.	
XX	TPO-mimetic peptide	
XX	TPO-mimetic peptide	
KW	Modified peptide; therapeutic agent; fusion; Fc domain; cancer;	
KW	autoimmune disease; cytostatic; antitumor; thrombolytic; VEGF;	
KW	immunopressure; EPO; TPO; CT1A4; mimetic; IL-1; TNF; antagonist;	
KW	MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;	
KW	cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;	
KW	vascular endothelial growth factor; matrix metalloproteinase;	
KW	asrama; thrombosis; pharmaceutical.	
XX	Synthetic.	
OS	Synthetic.	
PN	WO200024782-A2.	
PD	04-MAY-2000.	
XX	04-MAY-2000.	
XX	04-MAY-2000.	
FE	25-OCT-1999; 99WO-US25044.	
XX	25-OCT-1999; 99WO-US25044.	
PR	23-OCT-1998; 98US-0105371.	
XX	23-OCT-1998; 98US-0105371.	
PR	22-OCT-1999; 99US-0428082.	
XX	22-OCT-1999; 99US-0428082.	
PA	(AMGE-) AMGEN INC.	
XX	(AMGE-) AMGEN INC.	
XX	(AMGE-) AMGEN INC.	
PI	Feige U, Liu C, Cheetham J, Boone TC;	
XX	Feige U, Liu C, Cheetham J, Boone TC;	
XX	WPI; 2000-350702/30.	

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 PS Example 1; Page 190; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)-a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)-c-P1, -(L1)-c-P1-(L2)-d-P2,
 CC -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3, or -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3-(L4)-f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antitumoral, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AA69443
 CC to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA:
 SQ

Query Match 100.0%; Score 194; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 1.7e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLARAGGGGGGIEPTLRQWLARA 36
 Db 1 IEPTLRQWLARAGGGGGGIEPTLRQWLARA 36

RESULT 2
 AAB17293
 ID AAB17293 standard; Peptide: 36 AA.

XX AAB17293;
 AC
 XX
 DT 31-OCT-2000 (first entry)
 DE TPO-mimetic peptide sequence SEQ ID NO:349.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antitumoral; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumor necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.
 OS
 XX
 PN WO200024782-A2.
 PD 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.
 PF
 XX 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Felge U, Liu C, Cheetham J, Boone TC;
 PR WPI: 2000-350702/30.
 DR
 XX Novel composition of matter comprising an Fc domain and

PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 PS Example 1; Page 318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)-c-P1, -(L1)-c-P1-(L2)-d-P2,
 CC -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3, or -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3-(L4)-f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antitumoral, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AA69443
 CC to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA:
 SQ

Query Match 100.0%; Score 194; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 1.7e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLARAGGGGGGIEPTLRQWLARA 36
 Db 1 IEPTLRQWLARAGGGGGGIEPTLRQWLARA 36

RESULT 3
 AAY96525
 ID AAY96525 standard; peptide: 36 AA.

XX AAY96525;
 AC
 XX
 DT 04-SEP-2000 (first entry)
 DE Thrombopoietin mimetic peptide compound 6.

XX Thrombopoietin; mimetic; TMP; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker.

XX Synthetic.
 OS
 XX
 PN WO200024770-A2.
 PD 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.
 PF
 XX 23-OCT-1998; 98US-0105348.
 PR
 XX (AMGE-) AMGEN INC.
 XX

XX Key Location/Qualifiers
 FH Modified-site 1 /note="optionally linked to an Fc molecule"
 FT Peptide 1.14 /label= TMP_1
 FT Peptide 15.18 /label= linker
 FT Peptide 19.32 /label= TMP_2
 FT Modified-site 32 /note="optionally linked to an Fc molecule"

PI Liu C, Feige U, Cheetham J;
 XX
 XX WPI; 2000-365108/31.
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 PS
 PS Claim 16; Page 62; 91pp; English.
 XX
 XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L-1)-nTMP-2],
 CC is new, TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
 CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N,
 CC or E; X-9 = W, Y or F; X-1-0 = L, I, V, A, F, M, or K; X-1-1 = A, I, V,
 CC L, F, S, T, K, H, or E; X-1-2 = A, I, V, L, F, G, S, or Q; X-1-3 = R, K,
 CC T, V, N, Q or G; X-1-4 = A, I, V, L, F, T, R, E, or G; L-1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 XX Sequence 36 AA;
 SQ
 Query Match 100.0%; Score 194; DB 21; Length 36;
 Best Local Similarity 100.0%; Pred. No. 1.7e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 36
 DB 1 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 36
 RESULT 4
 AAY96528 standard; peptide; 41 AA.
 XX
 AC AAY96528;
 XX
 DT 04-SEP-2000 (first entry)
 XX
 DE Thrombopoietin mimetic peptide compound 9.
 XX
 DE Thrombopoietin mimetic peptide compound 9.
 XX
 KM Thrombopoietin mimetic; TMP; TPO; platelet; megakaryocyte; production;
 KM anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KM immunosuppressive; anti-inflammatory; linker.
 XX
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT Modified-site 1
 FT Peptide 6..19
 FT Peptide /label= TMP_1
 FT Peptide 20..27
 FT Peptide /label= linker
 FT Peptide 28..41
 FT Peptide /label= TMP_2
 XX
 PN WO200024770-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 22-OCT-1999; 99WO-US24834.
 XX
 PR 23-OCT-1998; 98US-0105348.
 XX

PA (AMGE-) AMGEN INC.
 XX
 XX Liu C, Feige U, Cheetham J;
 XX
 XX WPI; 2000-365108/31.
 XX
 PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia
 PS
 PS Claim 16; Page 65; 91pp; English.
 XX
 XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker [TMP-1-(L-1)-nTMP-2],
 CC is new, TMP-1 and TMP-2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
 CC X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and
 CC X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A;
 CC X-4 = P; X-5 = T or S; X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N,
 CC or E; X-9 = W, Y or F; X-1-0 = L, I, V, A, F, M, or K; X-1-1 = A, I, V,
 CC L, F, S, T, K, H, or E; X-1-2 = A, I, V, L, F, G, S, or Q; X-1-3 = R, K,
 CC T, V, N, Q or G; X-1-4 = A, I, V, L, F, T, R, E, or G; L-1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.
 XX
 XX Sequence 41 AA;
 SQ
 Query Match 100.0%; Score 194; DB 21; Length 41;
 Best Local Similarity 100.0%; Pred. No. 1.9e-16;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 36
 DB 6 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 41
 RESULT 5
 AAB17281
 ID AAB17281 standard; peptide; 42 AA.
 XX
 AC AAB17281;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:337.
 XX
 DE Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KM autoimmune disease; cytostatic; antiproliferative; thrombolytic; VEGF;
 KM immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KM MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KM cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KM vascular endothelial growth factor; matrix metalloproteinase;
 KM asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 XX
 PN WO200024782-A2.
 XX
 PD 04-MAY-2000.
 XX
 PF 25-OCT-1999; 99WO-US25044.
 XX
 PR 23-OCT-1998; 98US-0105371.
 XX
 PR 22-OCT-1999; 99US-0428082.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;

XX		WPI: 2000-350702/30.
DR		
XX		
PX		Novel composition of matter comprising an Fc domain and
PT		pharmacologically active peptides, useful for treating cancer and
PT		autoimmune diseases -
XX		
PS		Disclosure; Page 313; 608pp; English.
CC		
CC		The present invention describes composition of matter (I) comprising an
CC		Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC		(X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC		independently selected from -(L1)-c-P1, -(L1)c-P1-(L2)d-P2,
CC		-(L1)-c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC		where P1, P2, P3, and P4 = are each independently sequences of
CC		pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC		independently linkers; and a, b, c, d, e, and f = are each independently
CC		0 or 1, provided that at least 1 of a and b is 1. The composition can
CC		have cytosstatic, antilastmatic, thrombolytic and immunosuppressive
CC		activities. DNAs, vectors and host cells from the present invention can
CC		be used for producing pharmaceutical compositions. The compositions are
CC		useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC		The use of an Fc domain (rather than a rab domain) can provide a longer
CC		half-life or incorporate functions such as Fc receptor binding, protein
CC		A binding, complement fixation, and possibly placental transfer. AA69443
CC		to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC		sequences used in the exemplification of the present invention.
SQ		
	Sequence	42 AA;
	Query Match	100.0%; Score 194; DB 21; Length 42;
	Best Local Similarity	100.0%; Pred. No. 2e-16;
	Matches	36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 IEGPLRLRWLAARAGGGGIEGPLRLRWLAARA 36 	
Dd	7 IEGPTRLRWLAARAGGGGIEGPLRLRWLAARA 42	
RESULT 6		
AABI7282		
ID	AABI7282 standard; Peptide; 42 AA.	
AC		
AA	AABI7282;	
DT		
DE	31-OCT-2000 (first entry)	
DE	TPO-mimetic peptide sequence SEQ ID NO:338.	
KW	Modified peptide; therapeutic agent; fusion; Fc domain; cancer; autoimmune disease; cytostatic; antilastmatic; thrombolytic; VEGF; immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist; MMF; inhibitor; erythropoietin; thrombopoietin; interleukin 1; cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor; vascular endothelial growth factor; matrix metalloproteinase; asthma; thrombosis; pharmaceutical.	
OS	Synthetic.	
PN	WO200024782-A2.	
PD	04-MAY-2000.	
PF	25-OCT-1999; 99WO-US25044.	
PR	23-OCT-1998; 98US-0105371.	
PA	22-OCT-1999; 99US-0428082.	
AMGE-	- AMGEN INC.	
Feige U, Liu C, Cheetham J, Boone TC;		
WPI: 2000-350702/30.		

XX		Novel composition of matter comprising an Fc domain and
PT		pharmacologically active peptides, useful for treating cancer and
PT		autoimmune diseases -
XX		
PS		Disclosure; Page 313; 608bp; English.
XX		
CC	The present invention describes composition of matter (1) comprising an	
CC	Fc domain, pharmacologically active peptides, and linkers. Where (1) is:	
CC	(X1)-a-P1-(X2)b, where: P1 = an Fc domain; X1 and X2 = are each	
CC	independently selected from -(L1)-c-P1, -(L1)c-P1-(L2)d-P2,	
CC	-(L1)-c-P1-(L2)-d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4	
CC	where P1, P2, P3, and P4 = are each independently sequences of	
CC	pharmacologically active peptides; L1, L2, L3, and L4 = are each	
CC	independently linkers; and a, b, c, d, e, and f = are each independently	
CC	0 or 1, provided that at least 1 of a and b is 1. The composition can	
CC	have cytostatic, antitumorigenic, thrombolytic and immunosuppressive	
CC	activities. DNAs, vectors and host cells from the present invention can	
CC	be used for producing pharmaceutical compositions. The compositions are	
CC	useful for treating cancer, asthma, thrombosis, or autoimmune diseases.	
CC	The use of an Fc domain (rather than a Fab domain) can provide a longer	
CC	half-life or incorporate functions such as Fc receptor binding, protein	
CC	A binding, complement fixation, and possibly placental transfer. AA69443	
CC	to AA69526 and AAB1655 to AAB1803 represent nucleotide and amino acid	
CC	sequences used in the exemplification of the present invention.	
XX		
SQ	Sequence 42 AA;	
	Query Match 100.0%; Score 194; DB 21; Length 42;	
	Best Local Similarity 100.0%; Pred. No. 2e-16;	
	Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	1 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLARA 36	
Db	1 IEGPTLRQWLAARAGGGGGGIEGPTLRQWLARA 36	
RESULT 7		
AAB17308		
ID	AAB17308 standard; Peptide; 42 AA.	
XX		
AC	AAB17308;	
XX		
DT	31-OCT-2000 (first entry)	
XX		
DE	Synthetic TMP-TMP gene construction peptide SEQ ID NO:374.	
XX		
KM	Modified peptide; therapeutic agent; fusion; Fc domain; cancer;	
KM	autoimmune disease; cytostatic; antitumorigenic; thrombolytic; VEGF;	
KM	immunosuppressive; EPO; TPO; CM14; mimetic; IL-1; TNF; antagonist;	
KM	MMP; inhibitor; erythropoietin; thrombotic; interleukin 1;	
KM	Cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;	
KM	vascular endothelial growth factor; matrix metalloproteinase;	
KM	asthma; thrombosis; pharmaceutical.	
XX		
OS	Homo sapiens.	
OS	Synthetic.	
PX	WO200024782-A2.	
XX		
PD	04-MAY-2000.	
XX		
PF	25-OCT-1999; 99WO-US25044.	
XX		
PR	23-OCT-1998; 98US-0105371.	
XX		
PA	22-OCT-1999; 99US-0428082.	
XX		
PI	(AMGE-) AMGEN INC.	
XX		
PI	Feige U, Liu C, Cheetham J, Boone TC;	
XX		
WI	WPI; 2000-350702/30.	

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 327; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)-a-R1-(X2)-b, where: R1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)-c-P1, -(L1)-c-P1-(L2)-d-P2,
 CC -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3, or -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3-(L4)-f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antitumorigenic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AA69443
 CC to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

SO Sequence 42 AA:

Query Match 100.0%; Score 194; DB 21; Length 42;

Best Local Similarity 100.0%; Pred. No. 2e-16;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 36

DB 7 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 42

RESULT 8

AA96530

ID AAY96530 standard; Protein; 42 AA.

AC AAY96530;

DT 04-SEP-2000 (first entry)

DE Thrombopoietin mimetic peptide.

KW Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TMP; TPO; platelet;

KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;

KW anti-anaemic; dermatological; immunosuppressive; anti-inflammatory.

OS Synthetic.

PN WO200024770-A2.

PD 04-MAY-2000.

PF 22-OCT-1999; 99WO-US24834.

PR 23-OCT-1998; 98US-0105348.

PA (AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;

DR WPI; 2000-365108/31.

DR N-PSDB; AAA29225.

PT Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Example 2A; Page 48; 91pp; English.

CC Overlapping oligonucleotides were used to construct a synthetic
 CC gene encoding a thrombopoietin mimetic peptide (TMP), which
 CC was then fused in-frame to the Fc region of the human IgG1 chain (see
 CC AA96529). A compound which binds to an mpl receptor comprising a TMP
 CC dimer joined by a linker (TMP-1-(L1)-rTMP-2), is new. TMP-1 and TMP-2
 CC are amino acid sequences varying from at least 10 to 14 residues in
 CC length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2, X-2-X-1-3, X-2-X-1-4,
 CC X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and X-1-X-1-4. X-1 = I, A,
 CC V, L, S or R; X-2 = E, D, K or V; X-3 = G or A; X-4 = F; X-5 = T or S;
 CC X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N, or E; X-9 = W, Y or F;
 CC X-1-0 = L, I, V, A, F, M, or K; X-1-1 = A, I, V, L, F, S, T, K, H, or E;
 CC X-1-2 = A, I, V, L, F, G, S, or Q; X-1-3 = R, K, T, V, N, Q or G; X-1-4 =
 CC A, I, V, L, F, T, R, E, or G; L-1 = linker comprising 1 to 20 amino
 CC acids; and n = 0 or 1. The compounds bind to and activate the c-mpl
 CC receptor which mediates the activity of endogenous thrombopoietin. The
 CC TMPs are useful for increasing the production of platelets or platelet
 CC precursors (e.g. megakaryocytes) in a mammal, which is useful for
 CC treatment of diseases which involve thrombocytopenia, e.g. aplastic
 CC anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus
 CC associated ITP, and systemic lupus erythematosus.

SO Sequence 42 AA:

Query Match 100.0%; Score 194; DB 21; Length 42;

Best Local Similarity 100.0%; Pred. No. 2e-16;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 36

DB 7 IEPTLRQWLAAARAGGGGGGIEPTLRQWLAAAR 42

RESULT 9

AAB17311

ID AAB17311 standard; Peptide; 60 AA.

AC AAB17311;

DT 31-OCT-2000 (first entry)

DE Synthetic TMP-TMP-Fc gene construction peptide SEQ ID NO:385.

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;

KW autoimmune disease; cytostatic; antitumorigenic; thrombolytic; VEGF;

KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;

KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;

KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;

KW vascular endothelial growth factor; matrix metalloproteinase;

KW asthma; thrombosis; pharmaceutical.

OS Homo sapiens.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI; 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 331; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AA69443
CC to AA69526 and AA6955 to AA6955 to AA6955 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 60 AA:

Query Match 100.0%; Score 194; DB 21; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.8e-16;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 36
DB 2 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 37

RESULT 10
AA69560
ID AAB16960 standard; Protein: 269 AA.

XX AAB16960;

DT 31-OCT-2000 (first entry)

DE TMP-TMP-Fc protein sequence SEQ ID NO:10.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumor necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

XX Homo sapiens.
XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI: 2000-350702/30.

XX N-PSDB: AAA69446.

XX Novel composition of matter comprising an Fc domain and
XX pharmacologically active peptides, useful for treating cancer and
XX autoimmune diseases -
XX Example 2; Page 185-186; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AA69443
CC to AA69526 and AA6955 to AA6955 to AA6955 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 269 AA:

Query Match 100.0%; Score 194; DB 21; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.3e-15;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 36
DB 2 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 37

RESULT 11
AA69531
ID AAY96531 standard; Protein: 269 AA.

XX AAY96531;

DT 04-SEP-2000 (first entry)

DE Human IgG1 Fc TNP fusion protein.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TNP; platelet;
XX megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
XX anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

XX Homo sapiens.

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI: 2000-365108/31.

XX N-PSDB: AAA29229.

XX Thrombopoietic peptides which activate mpl receptors and increase the
XX production of platelets or platelet precursors, useful for treatment of
XX diseases which involve thrombocytopenia

XX Example 2A; Page 49-50; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
XX mimetic peptide (TMP) dimer joined by a linker (TMP-1-(L1)-TMP-2),
XX is new. TMP-1 and TMP-2 are amino acid sequences varying from at least
XX 10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2,
XX X-2-X-1-3, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and

CC X₁-X₁-X₁-X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or K; X₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, G, S, or Q; X₁₃ = R, K,
 CC T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

SO Sequence 269 AA;

Query Match 100.0%; Score 194; DB 21; Length 269;
 Best Local Similarity 100.0%; Pred. No. 1.3e-15;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAAAR 36
 ||||||||||||||||||||||||||||
 Db 234 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAAAR 269

RESULT 12
 AAB16959 ID AAB16959 standard; Protein; 268 AA.
 AC AAB16959;
 XX 31-OCT-2000 (first entry)
 DT
 DE Fe-TMP-TMP protein sequence SEQ ID NO:8.
 XX
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KM autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KM immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KM MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KM cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KM vascular endothelial growth factor; matrix metalloproteinase;
 KM asthma; thrombosis; pharmaceutical.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS WO200024782-A2.
 PN
 XX 04-MAY-2000.
 PD
 XX 25-OCT-1999; 99WO-US25044.
 PF
 XX 23-OCT-1998; 98US-0105371.
 PR
 XX 22-OCT-1999; 99US-0428082.
 PA
 XX (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 DR WPI: 2000-350702/30.
 DR N-PSDB; AAA69445.
 DR
 XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX Example 2; Page 182-183; 608pp; English.
 PS
 XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

SO Sequence 268 AA;

Query Match 97.9%; Score 190; DB 21; Length 268;
 Best Local Similarity 100.0%; Pred. No. 3.8e-15;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAAAR 35
 ||||||||||||||||||||||||
 Db 234 IEPTLRQWLAAARAGGGGGIEGPTLRQWLAAAR 268

RESULT 13
 AAB17301 ID AAB17301 standard; Peptide; 36 AA.
 AC AAB17301;
 XX 31-OCT-2000 (first entry)
 DT
 DE TPO-mimetic peptide sequence SEQ ID NO:357.
 XX
 XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KM autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KM immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KM MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KM cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KM vascular endothelial growth factor; matrix metalloproteinase;
 KM asthma; thrombosis; pharmaceutical.
 XX
 OS Synthetic.
 OS WO200024782-A2.
 PN
 XX 04-MAY-2000.
 PD
 XX 25-OCT-1999; 99WO-US25044.
 PF
 XX 23-OCT-1998; 98US-0105371.
 PR
 XX 22-OCT-1999; 99US-0428082.
 PA
 XX (AMGE-) AMGEN INC.
 XX
 PI Feige U, Liu C, Cheetham J, Boone TC;
 DR WPI: 2000-350702/30.
 DR
 XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -
 XX
 XX Example 1; Page 321; 608pp; English.
 PS
 XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAB69443
 CC to AAB69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 36 AA:

Query Match 95.9%; Score 186; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 1.5e-15;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 36
 1 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 36

RESULT 14
 AAY96523
 ID AAY96523 standard; peptide: 36 AA.

XX AAY96523:

DT 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 4.

XX Thrombopoietin mimetic; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KM immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14 /label= TMP_1

FT Peptide 15..22 /label= linker

FT Modified-site 18 /note= "optionally modified by bromoacetyl or PEG"

FT Peptide 23..36 /label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TMP) dimer joined by a linker (TMP_1-(L1)-TMP_2),
 CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

CC 10 to 14 residues in length comprising X₂-X₁0, X₂-X₁1, X₂-X₁2,
 CC X₂-X₁3, X₂-X₁4, X₁-X₁0, X₁-X₁1, X₁-X₁2, X₁-X₁3, and
 CC X₁-X₁4. X₁ = I, A, V, L, S or R; X₂ = E, D, K or V; X₃ = G or A;
 CC X₄ = P; X₅ = T or S; X₆ = L, I, V, A or F; X₇ = R or K; X₈ = Q, N,
 CC or E; X₉ = W, Y or F; X₁₀ = L, I, V, A, F, M, or R; X₁₁ = A, I, V,
 CC L, F, S, T, K, H, or E; X₁₂ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC T, V, N, Q or G; X₁₄ = A, I, V, L, F, T, R, E, or G; L₁ = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TMPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA:

Query Match 95.9%; Score 186; DB 21; Length 36;
 Best Local Similarity 97.2%; Pred. No. 1.5e-15;

Matches 35; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 36
 1 IEPTLRQWLARAGGGGGIEGPTLRQWLARA 36

RESULT 15
 AAB17303
 ID AAB17303 standard; peptide: 36 AA.

XX AAB17303:

DT 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:359.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 322; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X₁)-a-F₁-(X₂)_b, where: F₁ = an Fc domain; X₁ and X₂ = are each
 CC independently selected from -(L₁)-c-P₁-(L₂)-d-P₂,
 CC -(L₁)-c-P₁-(L₂)-d-P₂-(L₃)-e-P₃, or -(L₁)-c-P₁-(L₂)-d-P₂-(L₃)-e-P₃-(L₄)-f-P₄
 CC where P₁, P₂, P₃, and P₄ = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antitumoral, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AA69443
CC to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 36 AA:

Query Match 95.4%; Score 185; DB 21; Length 36;
Best Local Similarity 97.2%; Pred. No. 2e-15; Mismatches 0; Gaps 0;
Matches 35; Conservative 0; Indels 1; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 36
Db 1 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 36
|||||
AAB17307 standard; Peptide: 36 AA.
AAB17307;
31-OCT-2000 (first entry)
TPO-mimetic peptide sequence SEQ ID NO:363.
DE TPO-mimetic peptide sequence SEQ ID NO:363.
XX
DE Modified peptide: therapeutic agent; fusion: Fc domain; cancer;
KW autoimmune disease; cytostatic; antitumoral; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; Interleukin 1;
KW cyclooxycase T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
XX 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheatham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Example 1; Page 324; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-P1-(X2)b, where: P1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antitumoral, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AA69443
CC to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 36 AA:

Query Match 95.4%; Score 185; DB 21; Length 36;
Best Local Similarity 97.2%; Pred. No. 2e-15; Mismatches 0; Gaps 0;
Matches 35; Conservative 0; Indels 1; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 36
Db 1 IEPTLRQWLAAAGGGGGGIEPTLRQWLAAARA 36
|||||
AAY96524 standard; peptide: 36 AA.
AAY96524;
04-SEP-2000 (first entry)
Thrombopoietin mimetic peptide compound 5.
XX
DE Thrombopoietin mimetic peptide compound 5.
XX
KW Thrombopoietin; mimetic; TWP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FH Modified-site 1
FT Peptide /note= "optionally linked to an Fc molecule"
FT Disulfide-bond 1,14 /label= TWP_1
FT 9..31
FT Peptide /note= "optional"
FT 15..22 /label= linker
FT 23..36 /label= TWP_2
FT Peptide /label= TWP_2
XX
PN WO200024770-A2.
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24834.
XX
PR 23-OCT-1998; 98US-0105348.
XX
PA (AMGE-) AMGEN INC.
XX
PI Liu C, Feige U, Cheatham J;
XX
DR WPI; 2000-365108/31.
XX
PT Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thrombocytopenia
XX
PS Claim 16; Page 62; 91pp; English.
XX
CC A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TWP) dimer joined by a linker [TWP_1-(L1)-nTWP_2],
CC is new. TWP_1 and TWP_2 are amino acid sequences varying from at least

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AA69443
CC to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 38 AA:
Query Match 94.3%; Score 183; DB 21; Length 38;
Best Local Similarity 94.7%; Pred. No. 3.6e-15;
Matches 36; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY 1 IEGPTLRQWLARAGG--GGGGGIEGPTLRQWLARA 36
Db 1 IEGPTLRQWLARAGGCGGGGIEGPTLRQWLARA 38
RESULT 20
AAB17304
ID AAB17304 standard; Peptide: 39 AA.
AC AAB17304;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:360.
XX
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthina; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
XX
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
XX Example 1; Page 323; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AA69443
CC to AA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 39 AA:
Query Match 94.1%; Score 182.5; DB 21; Length 39;
Best Local Similarity 92.3%; Pred. No. 4.2e-15;
Matches 36; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
QY 1 IEGPTLRQWLARAGG--GGGGGIEGPTLRQWLARA 36
Db 1 IEGPTLRQWLARAGGKREGGGGIEGPTLRQWLARA 39
RESULT 21
AAB17305
ID AAB17305 standard; Peptide: 39 AA.
AC AAB17305;
XX
DT 31-OCT-2000 (first entry)
XX
DE TPO-mimetic peptide sequence SEQ ID NO:361.
XX
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthina; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
XX
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
XX Example 1; Page 323; 608pp; English.
XX
CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are

activate the c-Mol receptor which mediates the activity of endogenous thrombopoietin. The TMs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

Sequence 36 AA:

Query Match 93.8%; Score 182; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 4,5e-15;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36

RESULT 24

AB17296
ID AAB17296 standard; Peptide; 42 AA.

AC AAB17296;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:352.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)-a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)-c-P1, -(L1)-c-P1-(L2)-d-P2, -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3, or -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3-(L4)-f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer

half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Sequence 42 AA:

Query Match 93.3%; Score 181; DB 21; Length 42;
Best Local Similarity 85.7%; Pred. No. 6,9e-15;
Matches 36; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

OY 1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 36
1 IEGPTLRQWLAAARAGGGGGGIEGPTLRQWLAAARA 42

RESULT 25

AB17292
ID AAB17292 standard; Peptide; 35 AA.

AC AAB17292;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:348.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
XX immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
XX MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
XX cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
XX vascular endothelial growth factor; matrix metalloproteinase;
XX asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

WPI; 2000-350702/30.

Novel composition of matter comprising an Fc domain and pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

Example 1; Page 317-318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)-a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)-c-P1, -(L1)-c-P1-(L2)-d-P2, -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3, or -(L1)-c-P1-(L2)-d-P2-(L3)-e-P3-(L4)-f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 35 AA;

Query Match 91.5%; Score 177.5; DB 21; Length 35;
Best Local Similarity 97.2%; Pred. No. 1.5e-14;
Matches 35; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy 1 IEPTLRQWLAAARAGGGGIEPTLRQWLAAAR 36
Db 1 IEPTLRQWLAAARAGGGGIEPTLRQWLAAAR 35

RESULT 26
AAB17302

ID AAB17302 standard; Peptide: 40 AA.

AC AAB17302;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:358.

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antilasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI; 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

PS Example 1; Page 322; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antilasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 40 AA;

Query Match 89.7%; Score 174; DB 21; Length 40;
Best Local Similarity 87.5%; Pred. No. 4.5e-14;
Matches 35; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

Oy 1 IEPTLRQWLAAARAGGGG---GGGIEPTLRQWLAAAR 36
Db 1 IEPTLRQWLAAARAGGGKBRACGGGIEPTLRQWLAAAR 40

RESULT 27
AAB17291

ID AAB17291 standard; Peptide: 34 AA.

AC AAB17291;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:347.

KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antilasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

OS Synthetic.

PN WO200024782-A2.

PD 04-MAY-2000.

PF 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI; 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

PS Example 1; Page 317; 608pp; English.

CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antilasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX Sequence 34 AA;

Query Match 88.1%; Score 171; DB 21; Length 34;
Best Local Similarity 94.4%; Pred. No. 8.6e-14;
Matches 34; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

OY 1 IEPTLRQWLARAGGGGGIEPTLRQWLARA 36
1 IEPTLRQWLARA--GGGGGIEPTLRQWLARA 34

RESULT 28

AAB17298
ID AAB17298 standard; Peptide; 36 AA.

AC AAB17298;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:354.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

OS WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI, 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

PS Example 1; Page 320; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAB69443
CC A binding, complement fixation, and possibly placental transfer. AAB69443
CC sequences used in the exemplification of the present invention.

SO Sequence 36 AA;

Query Match 86.6%; Score 168; DB 21; Length 36;

Best Local Similarity 94.4%; Pred. No. 2.1e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 IEPTLRQWLARAGGGGGIEPTLRQWLARA 36
1 IEPTLRQWLARA--GGGGGIEPTLRQWLARA 36

RESULT 29

AAB17299
ID AAB17299 standard; Peptide; 36 AA.

AC AAB17299;

DT 31-OCT-2000 (first entry)

DE TPO-mimetic peptide sequence SEQ ID NO:355.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
XX autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

OS WO200024782-A2.

PN 04-MAY-2000.

PD 25-OCT-1999; 99WO-US25044.

PR 23-OCT-1998; 98US-0105371.

PR 22-OCT-1999; 99US-0428082.

PA (AMGE-) AMGEN INC.

PI Feige U, Liu C, Cheetham J, Boone TC;

DR WPI, 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

PS Example 1; Page 320-321; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
XX Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAB69443
CC A binding, complement fixation, and possibly placental transfer. AAB69443
CC sequences used in the exemplification of the present invention.

SO Sequence 36 AA;

Query Match 86.6%; Score 168; DB 21; Length 36;

Best Local Similarity 94.4%; Pred. No. 2.1e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARAGGGGGIEPTLRQWLARA 36
 ||||||| ||||||| ||||||| ||||||| |||||
 Db 1 IEPTLRQWLARAGGGGGIEPTLRQWLARA 36

RESULT 30

AA96521
 ID AAY96521 standard; peptide: 36 AA.

AC AAY96521;

DT 04-SEP-2000 (first entry)

DE Cyclic or linear thrombopoietin mimetic peptide compound 2.

XX Thrombopoietin; mimetic; TWP; TPO; platelet; megakaryocyte; production;
 KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
 KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note="optionally linked to an Fc molecule"

FT Peptide 1..14 /label= TWP_1

FT Disulfide-bond 9..31 /note="optional"

FT Peptide 15..22 /label= linker

FT Peptide 23..36 /label= TWP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI: 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

PS Claim 16; Page 61; 91pp: English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (TWP) dimer joined by a linker [TWP_1-(L_1)-TWP_2],
 CC is new. TWP_1 and TWP_2 are amino acid sequences varying from at least
 CC 10 to 14 residues in length comprising X_2-X_1_0, X_2-X_1_1, X_2-X_1_2,
 CC X_2-X_1_3, X_2-X_1_4, X_1-X_1_0, X_1-X_1_1, X_1-X_1_2, X_1-X_1_3, and
 CC X_1-X_1_4. X_1 = I, A, V, L, S or R; X_2 = E, D, K or V; X_3 = G or A;
 CC X_4 = P; X_5 = T or S; X_6 = L, I, V, A or F; X_7 = R or K; X_8 = Q, N,
 CC or E; X_9 = W, Y or F; X_1_0 = L, I, V, A, F, M, or K; X_1_1 = A, I, V,
 CC L, F, S, T, K, H, or E; X_1_2 = A, I, V, L, F, G, S, or Q; X_1_3 = R, K,
 CC T, V, N, Q or G; X_1_4 = A, I, V, L, F, T, R, E, or G; L_1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TWPs are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

Query Match 86.6%; Score 168; DB 21; Length 36;
 Best Local Similarity 94.4%; Pred. No. 2.1e-13;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARAGGGGGIEPTLRQWLARA 36
 ||||||| ||||||| ||||||| ||||||| |||||
 Db 1 IEPTLRQWLARAGGGGGIEPTLRQWLARA 36

Search completed: October 9, 2002, 08:58:56
 Job time : 17.1874 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:55:27 ; Search time 5.98595 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838C-26

Perfect score: 1 IE9PTLRQWLARAGGGGGGIE9PTLRQWLARA 36

Sequence: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*

2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*

3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*

4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*

5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*

6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76.5	39.4	25	2	US-08-764-640-231
2	76.5	39.4	25	3	US-09-244-298A-231
3	76.5	39.4	25	4	US-09-516-704-231
4	73	37.6	14	2	US-08-764-640-13
5	73	37.6	14	2	US-08-764-640-193
6	73	37.6	14	3	US-08-973-225-13
7	73	37.6	14	3	US-08-973-225-193
8	73	37.6	14	3	US-09-244-298A-13
9	73	37.6	14	3	US-09-244-298A-193
10	73	37.6	14	4	US-09-516-704-13
11	73	37.6	14	4	US-09-516-704-193
12	73	37.6	15	2	US-08-764-640-17
13	73	37.6	15	2	US-08-764-640-185
14	73	37.6	15	3	US-08-973-225-17
15	73	37.6	15	3	US-08-973-225-185
16	73	37.6	15	3	US-09-244-298A-17
17	73	37.6	15	3	US-09-244-298A-185
18	73	37.6	15	4	US-09-516-704-17
19	73	37.6	15	4	US-09-516-704-185
20	73	37.6	16	2	US-08-764-640-18
21	73	37.6	16	2	US-08-764-640-194
22	73	37.6	16	2	US-08-764-640-232
23	73	37.6	16	3	US-08-973-225-18
24	73	37.6	16	3	US-08-973-225-194
25	73	37.6	16	3	US-08-973-225-220
26	73	37.6	16	3	US-09-244-298A-18
27	73	37.6	16	3	US-09-244-298A-194

28	73	37.6	16	3	US-09-244-298A-232	Sequence 232, App
29	73	37.6	16	4	US-09-516-704-18	Sequence 18, App1
30	73	37.6	16	4	US-09-516-704-194	Sequence 194, App
31	73	37.6	16	4	US-09-516-704-232	Sequence 232, App
32	69	35.6	14	2	US-08-764-640-195	Sequence 195, App
33	69	35.6	14	2	US-08-764-640-199	Sequence 199, App
34	69	35.6	14	3	US-08-973-225-195	Sequence 195, App
35	69	35.6	14	3	US-08-973-225-199	Sequence 199, App
36	69	35.6	14	3	US-09-244-298A-195	Sequence 195, App
37	69	35.6	14	3	US-09-244-298A-199	Sequence 199, App
38	69	35.6	14	4	US-09-516-704-195	Sequence 195, App
39	69	35.6	14	4	US-09-516-704-199	Sequence 199, App
40	69	35.6	15	2	US-08-764-640-196	Sequence 196, App
41	69	35.6	15	2	US-08-764-640-200	Sequence 200, App
42	69	35.6	15	2	US-08-764-640-209	Sequence 209, App
43	69	35.6	15	2	US-08-764-640-215	Sequence 215, App
44	69	35.6	15	3	US-08-973-225-196	Sequence 196, App
45	69	35.6	15	3	US-08-973-225-200	Sequence 200, App

ALIGNMENTS

RESULT 1
US-08-764-640-231
; Sequence 231, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dowder, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprence, Randolph B.
; APPLICANT: Poddaturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764, 640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36, 392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 231:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:

NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /product= "Ava"
US-08-764-640-231

Query Match 39.4%; Score 76.5; DB 2; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0058;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

OY 2 EGPTRLRLAARAGGGGGGIEGPTLRQWLA 33
DB 2 DGPTLRWISFXA-----DGPTLRWIS 24

RESULT 2
US-09-244-298A-231
Sequence 231, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/244, 298A

APPLICATION NUMBER: US/09/244, 298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids
TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: peptide
FEATURE:

NAME/KEY: Modified-site
LOCATION: 13

OTHER INFORMATION: /product= "Ava"
US-09-244-298A-231

Query Match 39.4%; Score 76.5; DB 3; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0058;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

OY 2 EGPTRLRLAARAGGGGGGIEGPTLRQWLA 33

DB 2 DGPTLRWISFXA-----DGPTLRWIS 24

RESULT 3
US-09-516-704-231
Sequence 231, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Poddaturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/516, 704

APPLICATION NUMBER: US/09/516, 704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 231:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids
TYPE: amino acid

STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: peptide
FEATURE:

NAME/KEY: Modified-site
LOCATION: 13

OTHER INFORMATION: /product= "Ava"
SEQUENCE DESCRIPTION: SEQ ID NO: 231:

US-09-516-704-231

Query Match 39.4%; Score 76.5; DB 4; Length 25;
Best Local Similarity 40.6%; Pred. No. 0.0058;
Matches 13; Conservative 8; Mismatches 2; Indels 9; Gaps 1;

OY 2 EGPTRLRLAARAGGGGGGIEGPTLRQWLA 33
DB 2 DGPTLRWISFXA-----DGPTLRWIS 24

RESULT 4
US-08-764-640-13
Sequence 13, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palanlappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-13

Query Match 37.6%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPLRLQWLARA 14
DB 1 IEGPLRLQWLARA 14

RESULT 5
US-08-764-640-193
Sequence 193, Application US/08764640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palanlappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirnce, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193

Query Match 37.6%; Score 73; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPLRLQWLARA 14
DB 1 IEGPLRLQWLARA 14

RESULT 6
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherril S.
APPLICANT: Mattheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Magstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A

FILED DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13

Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
DB 1 IEPTLRQWLARA 14

RESULT 7
US-08-973-225-193
Sequence 193, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherill S.
Matheakis, Larry C.
Schatz, Peter J.
Magstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-08-973-225-193

Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
DB 1 IEPTLRQWLARA 14

RESULT 8
US-09-244-298A-13
Sequence 13, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Gates, Christian
Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-Dec-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
DB 1 IEPTLRQWLARA 14

RESULT 9
US-09-244-298A-193
Sequence 193, Application US/09244298A

Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Poddaturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrublec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-193
Query Match 37.6%; Score 73; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IEPTLRQWLARA 14
DB 1 IEPTLRQWLARA 14
RESULT 10
US-09-516-704-13
Sequence 13, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Poddaturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrublec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13
Query Match 37.6%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IEPTLRQWLARA 14
DB 1 IEPTLRQWLARA 14
RESULT 11
US-09-516-704-193
Sequence 193, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprince, Randolph B.
APPLICANT: Poddaturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubic, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match 37.6%; Score 73; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLROWLAARA 14
|||||
DB 1 IEPTLROWLAARA 14

RESULT 12
US-08-764-640-17

Sequence 17, Application US/08/64640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barlett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubic, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-17

Query Match 37.6%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLROWLAARA 14
|||||
DB 1 IEPTLROWLAARA 14

RESULT 13

US-08-764-640-185
Sequence 185, Application US/08/64640
Patent No. 5869451
Patent No. 5869451 5837683
GENERAL INFORMATION:

APPLICANT: Dower, William J.
APPLICANT: Barlett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depince, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubic, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-185

Query Match 37.6%; Score 73; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLROWLAARA 14
|||||
DB 2 IEPTLROWLAARA 15

RESULT 14
US-08-973-225-17
Sequence 17, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haseiden, Sherill S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Nicholas R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-973-225-17
Query Match 37.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
DB 1 IEPTLRQWLARA 14

RESULT 15
US-08-973-225-185
Sequence 185, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haseiden, Sherill S.
Matheakis, Larry C.
Schatz, Peter J.

Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185
Query Match 37.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
DB 2 IEPTLRQWLARA 15

RESULT 16
US-09-244-298A-17
Sequence 17, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Haseiden, Richard W.
Deprince, Randolph B.
Poddurti, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17

Query Match
Best Local Similarity 37.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
Db 1 IEPTLRQWLARA 14

RESULT 17
US-09-244-298A-185
Sequence 185, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depiince, Randolph B.
APPLICANT: Poddurti, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-185

Query Match
Best Local Similarity 37.6%; Score 73; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
Db 2 IEPTLRQWLARA 15

RESULT 18
US-09-516-704-17
Sequence 17, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Magstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depiince, Randolph B.
APPLICANT: Poddurti, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/516,704
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match
Best Local Similarity 37.6%; Score 73; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLARA 14
Db 1 IEPTLRQWLARA 14

Db 1 IEPTLRWLARA 14

RESULT 19

US-09-516-704-185

Sequence 185, Application US/09516704

Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwiria, Steven E.

Gates, Christian

Schatz, Peter J.

Balasubramanian, Palaniappan

Wagstrom, Christopher R.

Hendren, Richard W.

Depirnce, Randolph B.

Poddaturi, Surekha

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/516,704

FILING DATE: 01-Mar-2000

CLASSIFICATION: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Hrubic, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 185:

SEQUENCE CHARACTERISTICS:

LENGTH: 15 amino acids

TYPE: amino acid

STRANDEDNESS: <unknown>

MOLECULE TYPE: linear

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 185:

US-09-516-704-185

Query Match 37.6%, Score 73; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.0086;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRWLARA 14

Db 2 IEPTLRWLARA 15

RESULT 20

US-08-764-640-18

Sequence 18, Application US/08764640

Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

Cwiria, Steven E.

APPLICANT: Gates, Christian

APPLICANT: Schatz, Peter J.

APPLICANT: Balasubramanian, Palaniappan

APPLICANT: Wagstrom, Christopher R.

APPLICANT: Hendren, Richard W.

APPLICANT: Depirnce, Randolph B.

APPLICANT: Poddaturi, Surekha

APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome

STREET: Five Moore Drive, P.O. Box 13398

CITY: Research Triangle Park

STATE: NC

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640

FILING DATE: 11-DEC-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Hrubic, Robert T.

REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-248-1000

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 16 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Modified-site

LOCATION: 15

OTHER INFORMATION: /product="Beta-ala"

US-08-764-640-18

Query Match 37.6%, Score 73; DB 2; Length 16;

Best Local Similarity 100.0%; Pred. No. 0.0092;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRWLARA 14

Db 1 IEPTLRWLARA 14

RESULT 21

US-08-764-640-194

Sequence 194, Application US/08764640

Patent No. 5869451

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.

APPLICANT: Cwiria, Steven E.

APPLICANT: Gates, Christian

APPLICANT: Schatz, Peter J.

APPLICANT: Balasubramanian, Palaniappan

APPLICANT: Wagstrom, Christopher R.

APPLICANT: Hendren, Richard W.

APPLICANT: Depirnce, Randolph B.

APPLICANT: Poddaturi, Surekha

APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A RECEPTOR

NUMBER OF SEQUENCES: 244

CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-194

Query Match 37.6%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAAARA 14
|||||
Db 2 IEPTLRQWLAAARA 15

RESULT 22
US-08-764-640-232
Sequence 232, Application US/08764640
Patent No. 5869451
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depirince, Randolph B.
APPLICANT: Poddurti, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-232

Query Match 37.6%; Score 73; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEPTLRQWLAAARA 14
|||||
Db 2 IEPTLRQWLAAARA 15

RESULT 23
US-08-973-225-18
Sequence 18, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Duffin, David J.
APPLICANT: Gates, Christian
APPLICANT: Haselden, Sherril S.
APPLICANT: Matheakis, Larry C.
APPLICANT: Schatz, Peter J.
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide

FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product="beta-ala"
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-08-973-225-18

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. NO. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPLRLQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPLRLQWLAAARA 14

RESULT 24

US-08-973-225-194
Sequence 194, Application US/08973225A
Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESS: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:
NAME: Hrublec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-08-973-225-194

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. NO. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPLRLQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPLRLQWLAAARA 15

RESULT 25
US-08-973-225-220
Sequence 220, Application US/08973225A
Patent No. 6083913

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESS: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997

ATTORNEY/AGENT INFORMATION:
NAME: Hrublec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 220:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 220:
US-08-973-225-220

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. NO. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGPLRLQWLAAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPLRLQWLAAARA 15

RESULT 26
US-09-244-298A-18
Sequence 18, Application US/09244298A
Patent No. 6121238

GENERAL INFORMATION:

APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
APPLICANT: Dephrice, Randolph B.

APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "beta-ala"
US-09-244-298A-18

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 1 IEPTLRQWLAARA 14

RESULT 27
US-09-244-298A-194
Sequence 194, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher K.
APPLICANT: Hendren, Richard W.
APPLICANT: Depina, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC

COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-194

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEPTLRQWLAARA 14
Db 2 IEPTLRQWLAARA 15

RESULT 28
US-09-244-298A-232
Sequence 232, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwirla, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Depina, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392

REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 232:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-232

Query Match 37.6%; Score 73; DB 3; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLARA 14
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQWLARA 15

RESULT 29

US-09-516-704-18
Sequence 18, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Magstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

US-09-516-704-18

Query Match 37.6%; Score 73; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0092;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEPTLRQWLARA 14
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQWLARA 14

RESULT 30

US-09-516-704-194
Sequence 194, Application US/09516704
Patent No. 6251864

GENERAL INFORMATION:

APPLICANT: Dower, William J.

Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Magstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Podduri, Surekha
Podduri, Surekha

Search completed: October 9, 2002, 09:06:30
Job time: 5.98595 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 8.09368 Seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838c-26

Perfect score: 194

Sequence: 1 IEPTLRQWLARAGGGGGGIEGPTLRQWLARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	69	35.6	500	2	T20961	hypothetical prote
2	68.5	35.3	488	2	G87033	probable ATP/GTP-b
3	68.5	35.3	518	2	S72938	hflx protein - Myc
4	66.5	34.3	495	2	D70505	probable HflX - My
5	64	33.0	201	2	T45792	hypothetical prote
6	64	33.0	331	2	T26807	hypothetical prote
7	64	33.0	333	2	T26808	hypothetical prote
8	63.5	32.7	619	1	KSNCLT	laccase (EC 1.10.3
9	63.5	32.7	619	1	KSNCLT	laccase (EC 1.10.3
10	63	32.5	201	2	JQ1094	hypothetical 20.2K
11	63	32.5	490	2	T09084	phosphatidylinosit
12	62.5	32.2	209	2	B42687	neurotrophin-4 pre
13	61.5	31.7	487	2	B39490	subtilisin-like pr
14	61.5	31.7	652	1	JC2191	subtilisin-like pr
15	61.5	31.7	962	2	JC5571	subtilisin-like pr
16	61.5	31.7	969	1	A39490	subtilisin-like pr
17	61.5	31.7	975	2	JC5570	subtilisin-like pr
18	61	31.4	415	2	D95664	hypothetical prote
19	61	31.4	443	1	S93334	transcription fact
20	61	31.4	445	1	S13224	transcription fact
21	61	31.4	593	1	KRHUO	keratin 10, type I
22	61	31.4	777	2	S65543	3',5'-cyclic-nucle
23	61	31.4	1168	1	MMAXIC	myosin heavy chain
24	60.5	31.2	210	2	A42687	neurotrophin-4 pre
25	60.5	31.2	864	2	A48266	protein-tyrosine k
26	60	30.9	285	2	S93312	probable membrane
27	60	30.9	323	2	S20099	transforming prote
28	60	30.9	569	1	KRMSE1	keratin, 59K type
29	60	30.9	649	2	S58064	hdc protein - frui

30 60 30.9 806 2 T13690 hypothetical prote
31 60 30.9 888 2 T58378 tyrosine kinase -
32 60 30.9 962 2 T04124 receptor-like prot
33 60 30.9 1325 2 T13386 hypothetical prote
34 59.5 30.7 327 2 B84781 hypothetical prote
35 59.5 30.7 339 2 T06612 hypothetical prote
36 59.5 30.7 403 2 A53662 homeotic protein H
37 59.5 30.7 443 2 E96495 hypothetical prote
38 59.5 30.7 867 2 S57795 probable deoxyribo
39 59 30.4 80 2 T10550 hypothetical prote
40 59 30.4 199 2 T48099 hypothetical prote
41 59 30.4 250 2 H85067 hypothetical prote
42 59 30.4 270 2 T35365 hypothetical prote
43 59 30.4 346 1 S35500 hypothetical prote
44 59 30.4 367 2 JC6087 heterogeneous ribo
45 59 30.4 396 2 T49109 helix-loop-helix t
glycine-rich prote

ALIGNMENTS

RESULT 1

T20961

hypothetical protein F15B9.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T20961

R:Percy, C.

Submitted to the EMBL Data Library, August 1996

A:Reference number: Z19351

A:Accession: T20961

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-500 <WIL>

A:Cross-references: EMBL:Z78013; PIDN:CAB01420.1; GSPDB:GN00023; CESP:F15B9.5

A:Experimental source: clone F15B9

C:Genetics:

A:Gene: CESP:F15B9.5

A:Map position: 5

A:Introns: 46/3; 63/3; 125/2; 162/2; 283/3; 391/1; 446/1

Query Match 35.6%; Score 69; DB 2; Length 500;
Best Local Similarity 56.5%; Pred. No. 3.3;
Matches 13; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 GPTLRQWLARAGGGGGGIEG 25

Db 429 GSMLGRFLSNRGGGGGGMG 451

RESULT 2

G87033

probable ATP/GTP-binding protein [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: G87033

R:Coile, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: AB6909; MUID:21128732; PMID:11234002

A:Accession: G87033

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-488 <STO>

A:Cross-references: GB:AL450380; NID:g13093026; PIDN:CAC31378.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML0997

C:Superfamily: GTP-binding protein hflx; translation elongation factor Tu homology

Query Match 35.3%; Score 68.5; DB 2; Length 488;

```

Best Local Similarity 46.7%; Pred. No. 3.7;
Matches 14; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY      4 PTLRW-----LAARAGGGGGGGIEGP 26
      | | | | : | | | | | | | |
DDB      189 PRLRGWESMSRQVGGRAGGGGGVGLRGP 218

RESULT 3
S72938
hflx protein - Mycobacterium leprae
N:Alternate names: B2235_C2.202 protein
C:Species: Mycobacterium leprae
#seqid change 25-Nov-1997 #text change 23-Mar-2001

```

C:Accession: S72938
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid B2235.
A:Reference number: S72587
A:Accession: S72938
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-518 <SMI>
A:Cross-references: EMBL:U00019; NID:g467079; PIDN:RAAL7274.1; PID:g467091
C:Genetics:
A:Start codon: GTG
C:Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match	35.3%;	Score	68.5;	DB	2;	Length	518;
Best Local Similarity	46.7%;	Pred. No.	3.9;				
Matches	14;	Conservative	2;	Mismatches	7;	Indels	7;
						Gaps	1;

```

QY      4  PTLRW-----LAARAGGGGGGGIEGP 26
      | | | | : | | | | | | | |
Db      219  PRLRGESMSRVGGRRAGGGGVGLRGP 248

RESULT 4
D70505
probable Hflx - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 02-Sep-2000

```

R; Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Whitehead, S.; Brown, J. E.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Sgares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.

A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A: Reference number: A70500; MUID: 98295987

A: Accession: D70505

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-495 <COL>

A: Cross-references: GB:298209; GB:AL123456; NID:g3261838; PID:CAB10901.1; PID:e332282;

A: Experimental source: strain H37Rv

C: Genetics:

A: Gene: hflX

```

Query Match      34.3%   Score 66.5;   DB 2;   Length 495;
Best Local Similarity 46.7%   Pred. No. 6;
Matches 14;   Conservative 1;   Mismatches 8;   Indels 7;   Gaps 1;

Qy  4  PTLROW-----LAARAGGGGGGGIEGP 26
      ||| |          |||| ||| | : ||
Db 199 PRLRGWGESMSRQAGGRAGGGGGYGLRGP 228

RESULT 5
T49792

```

Query Match 33.0%; Score 64; DB 2; Length 333;
 Best Local Similarity 76.9%; Pred. No. 7.7;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27
 Db 169 GGGGGGGVGP 181

RESULT 8
 KSNCLT
 Laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)
 N/Alternate names: urishiol oxidase
 C/Species: Neurospora crassa
 C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
 C/Accession: A28523; A29762
 R/Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
 J. Biol. Chem. 263, 885-896, 1988
 A/Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and
 A/Reference number: A28523; MUID:88087214
 A/Accession: A28523
 A/Molecule type: DNA
 A/Residues: 1-619 <GER>
 A/Cross-references: EMBL:M14554
 R/Germann, U.A.; Lerch, K.
 Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986
 A/Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospora
 A/Reference number: A29762; MUID:87067412
 A/Accession: A29762
 A/Molecule type: DNA
 A/Residues: 379-619 <GE2>
 A/Cross-references: GB:M14554; NID:gl68823; PIDN:AAA33590.1; PID:gl68824
 C/Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone
 C/Genetics:
 A/Introns: 86/3
 C/Superfamily: laccase
 C/Keywords: copper; glycoprotein; oxidoreductase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-49/Domain: propeptide #status predicted <PRO>
 F:50-619/Product: laccase #status predicted <MAT>
 F:84-215/Domain: middle beta-barrel #status predicted <BB2>
 F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
 F:139,282,295,340,422,444/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:144,480/Binding site: copper (His) (type 2) #status predicted
 F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status pred
 F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 32.7%; Score 63.5; DB 1; Length 619;
 Best Local Similarity 57.7%; Pred. No. 15;
 Matches 15; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 7 ROWLAARAGGGGGGIEGPTLRQ-W 31
 Db 39 RQDSQAERYGGGGGCGNSPTNRQW 64

RESULT 10
 JQ1094
 Hypothetical 20.2K protein - tomato ringspot virus
 C/Species: tomato ringspot virus
 C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 08-Oct-1999
 C/Accession: JQ1094
 R/Rott, M.E.; Tremaine, J.H.; Rochon, D.M.
 J. Gen. Virol. 72, 1505-1514, 1991
 A/Title: Nucleotide sequence of tomato ringspot virus RNA-2.
 A/Reference number: JQ1093; MUID:91311402
 A/Accession: JQ1094
 A/Status: translation not shown
 A/Molecule type: genomic RNA
 A/Residues: 1-201 <ROT>
 A/Cross-references: GB:D12477; GB:D01129; NID:g222674; PIDN:BAA02044.1; PID:d1002526;
 A/Experimental source: strain raspberry

Query Match 32.5%; Score 63; DB 2; Length 201;
 Best Local Similarity 61.5%; Pred. No. 6.2;
 Matches 16; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 13 RAGGGGGGGIE---GPTLRQWLAA 34
 Db 13 RAGGGGGGGKEVFKAGRTLLKVIKA 38

RESULT 11
 T09084
 phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)
 C/Species: Chlamydomonas reinhardtii
 C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C/Accession: T09084
 R/Molendijk, A.J.; Irvine, R.F.
 Plant Mol. Biol. 37, 53-66, 1998
 A/Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylyno
 A/Reference number: Z16411; MUID:98281574
 A/Accession: T09084
 A/Status: preliminary;
 A/Molecule type: DNA
 A/Residues: 1-490 <MOL>
 A/Cross-references: EMBL:U97563; NID:g2109290; PIDN:AAC50018.1; PID:g2109291
 A/Experimental source: strain cw-15
 C/Genetics:
 A/Introns: 265/3; 331/3; 370/3; 455/1; 481/3

Query Match 32.5%; Score 63; DB 2; Length 490;
 Best Local Similarity 48.6%; Pred. No. 14;
 Matches 17; Conservative 2; Mismatches 6; Indels 10; Gaps 3;

Query Match 33.0%; Score 64; DB 2; Length 333;
 Best Local Similarity 76.9%; Pred. No. 7.7;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27
 Db 169 GGGGGGGVGP 181

RESULT 8
 KSNCLT
 Laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)
 N/Alternate names: urishiol oxidase
 C/Species: Neurospora crassa
 C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
 C/Accession: A28523; A29762
 R/Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
 J. Biol. Chem. 263, 885-896, 1988
 A/Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and
 A/Reference number: A28523; MUID:88087214
 A/Accession: A28523
 A/Molecule type: DNA
 A/Residues: 1-619 <GER>
 A/Cross-references: EMBL:M14554
 R/Germann, U.A.; Lerch, K.
 Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986
 A/Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospora
 A/Reference number: A29762; MUID:87067412
 A/Accession: A29762
 A/Molecule type: DNA
 A/Residues: 379-619 <GE2>
 A/Cross-references: GB:M14554; NID:gl68823; PIDN:AAA33590.1; PID:gl68824
 C/Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone
 C/Genetics:
 A/Introns: 86/3
 C/Superfamily: laccase
 C/Keywords: copper; glycoprotein; oxidoreductase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-49/Domain: propeptide #status predicted <PRO>
 F:50-619/Product: laccase #status predicted <MAT>
 F:84-215/Domain: middle beta-barrel #status predicted <BB2>
 F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
 F:139,282,295,340,422,444/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:144,480/Binding site: copper (His) (type 2) #status predicted
 F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status pred
 F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 32.7%; Score 63.5; DB 1; Length 619;
 Best Local Similarity 57.7%; Pred. No. 15;
 Matches 15; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 7 ROWLAARAGGGGGGIEGPTLRQ-W 31
 Db 39 RQDSQAERYGGGGGCGNSPTNRQW 64

RESULT 9
 KSNCLT
 Laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)
 N/Alternate names: urishiol oxidase
 C/Species: Neurospora crassa
 C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
 C/Accession: B28523
 R/Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
 J. Biol. Chem. 263, 885-896, 1988
 A/Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and
 A/Reference number: A28523; MUID:88087214
 A/Accession: B28523
 A/Molecule type: DNA
 A/Residues: 1-619 <GER>
 A/Cross-references: EMBL:M18334; NID:gl68827; PIDN:AAA33592.1; PID:gl68828
 C/Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone

Nucleic Acids Res. 21, 253-258, 1993
 A:Title: cDNA cloning of human N-Oct 3, a nervous-system specific POU domain transcript
 A:Reference number: S30296; MUID:93181199
 A:Accession: S30296
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-25, 'G', 27-443 <SCW>
 A:Cross-references: EMBL:Z11933
 A:Experimental source: tissue-type brain
 C:Genetics:
 A:Gene: GDB:POU3F2; ONF7
 A:Cross-references: GDB:22816; OMIM:600494
 A:Map position: 6q16-6q16
 C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology
 C:Keywords: alternative initiators; DNA binding; homeobox; nucleus; transcription regulation
 F:1-443/Product: transcription factor Brn-2 #status experimental <MAT1>
 F:68-90/Region: glycine-rich
 F:125-149/Region: glutamine-rich
 F:151-165/Region: histidine/proline-rich
 F:181-443/Product: transcription factor Oct-5a #status experimental <MAT2>
 F:200-443/Product: transcription factor Oct-5b #status experimental <MAT3>
 F:211-259/Region: histidine/proline-rich
 F:269-336/Domain: POU domain homology <POU>
 F:355-411/Domain: homeobox homology <HOX>

Query Match 31.4%; Score 61; DB 1; Length 443;
 Best Local Similarity 66.7%; Pred. No. 21;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 QWLAARAGGGGGGG 22
 II: I : IIIIIII
 Db 60 QWITALSHGGGGGG 74

RESULT 20
 S31224
 transcription factor Brn-2 - mouse
 N:Alternate names: class III POU domain protein brain-2
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 22-Jun-1999
 C:Accession: S31224
 R:Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.
 Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992
 A:Title: Structure and evolution of four POU domain genes expressed in mouse brain.
 A:Reference number: S31223; MUID:92228768
 A:Accession: S31224
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-445 <HAR>
 A:Cross-references: EMBL:M88300; NID:g200446; PIDN:AAA39961.1; PID:g200447
 C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:68-90/Region: glycine-rich
 F:125-151/Region: glutamine-rich
 F:153-165/Region: histidine/proline-rich
 F:213-261/Region: histidine/proline-rich
 F:271-338/Domain: POU domain homology <POU>
 F:357-413/Domain: homeobox homology <HOX>

Query Match 31.4%; Score 61; DB 1; Length 445;
 Best Local Similarity 66.7%; Pred. No. 21;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 QWLAARAGGGGGGG 22
 II: I : IIIIIII
 Db 60 QWITALSHGGGGGG 74

RESULT 21
 KRHD0
 keratin 10, type I, cytoskeletal - human
 N:Alternate names: cytokeratin 10
 C:Species: Homo sapiens (man)

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 10-Dec-1999
 C:Accession: S02158; C38182; B38182; PC1102; S14666; S14669
 R:Reier, M.; Franke, W.W.
 J. Mol. Biol. 204, 841-856, 1988
 A:Title: Identification of an orthologous mammalian cytochrome gene. High degree of
 A:Reference number: S02158; MUID:89125611
 A:Accession: S02158
 A:Molecule type: DNA
 A:Residues: 1-593 <RIE>
 A:Cross-references: EMBL:X14487; NID:g28316; PIDN:CAA32649.1; PID:g28317
 A:Experimental source: clone lambda-KH10-5
 R:Korge, B.P.; Gan, S.Q.; McBride, O.W.; Mischke, D.; Steinert, P.M.
 Proc. Natl. Acad. Sci. U.S.A. 89, 910-914, 1992
 A:Title: Extensive size polymorphism of the human keratin 10 chain resides in the C-
 A:Reference number: A38182; MUID:92141228
 A:Accession: C38182
 A>Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 452-593 <KOR1>
 A:Cross-references: PIDN:AA21315.1; PID:g244509
 A:Note: sequence extracted from NCBI backbone (NCBIP:79427)
 A:Accession: B38182
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 452-463, 'P', 465-507, 'Y', 523-593 <KOR2>
 A:Cross-references: PIDN:AA21314.1; PID:g244508
 A:Note: sequence extracted from NCBI backbone (NCBIP:79431)
 R:Tachenko, A.V.; Buchman, V.L.; Bliskovsky, V.V.; Shvets, Y.P.; Kisselev, L.L.
 Gene 116, 245-251, 1992
 A:Title: Exons I and VII of the gene (Ker10) encoding human keratin 10 undergo struct
 A:Reference number: PC1102; MUID:92339897
 A:Accession: PC1102
 A:Molecule type: mRNA
 A:Residues: 'G', 198-407, 'Q', 409-450, 'G', 452-486, 491-524, 534-593 <TKA>
 A:Cross-references: GB:M77663; NID:g186628; PIDN:AAA59199.1; PID:g186629
 A:Experimental source: embryonic skin, clone HK51
 R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.
 Mol. Biol. Rep. 12, 277-283, 1987
 A:Title: Sequence of a cDNA encoding human keratin No 10 selected according to struct
 A:Reference number: S14666; MUID:88122104
 A:Accession: S14666
 A:Molecule type: mRNA
 A:Residues: 130-278, 'YV', 281-311, 'I', 313-339, 'V', 341-373, 'R', 375-407, 'Q', 409-459, 'RS'
 56-579, 'P', 581-593 <DAR>
 A:Cross-references: EMBL:M19156; NID:g186769
 A:Note: the sequence from Fig. 3 is inconsistent with the nucleotide sequence from Fi
 R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.
 submitted to the EMBL Data Library, May 1988
 A:Reference number: S14667
 A:Accession: S14669
 A:Molecule type: mRNA
 A:Residues: 130-278, 'YV', 281-311, 'I', 313-339, 'V', 341-373, 'R', 375-407, 'Q', 409-459, 'RS'
 56-593 <DAR2>
 A:Cross-references: EMBL:M19156; NID:g186769; PIDN:AAA59468.1; PID:g307086
 A:Note: the translated sequence in GenBank entry HUMKRT10A, release 111.0, differs fr
 C:Genetics:
 A:Gene: GDB:KRT10; KPP
 A:Cross-references: GDB:118828; OMIM:148080
 A:Map position: 17q12-17q21
 A:Introns: 209/3; 237/2; 289/3; 343/3; 385/3; 458/2; 592/3
 A:Note: this gene encodes variants with considerable length polymorphism
 A:Note: mutations in this gene can cause epidermolytic hyperkeratosis and keratosis p
 C:Complex: heterotetramer of two type I and two type II proteins, usually keratin 1 (C
 C:Superfamily: cytoskeletal keratin
 C:Keywords: coiled coil; heterotetramer; intermediate filament; polymorphism
 F:1-145/Domain: head <HEA>
 F:146-456/Domain: helical rod #status predicted <ROD>
 F:457-593/Domain: tail <TAI>

Query Match 31.4%; Score 61; DB 1; Length 593;
 Best Local Similarity 52.6%; Pred. No. 27;
 Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 7 ROWLAARAGGGGGGGIEG 25
: : : : : ||||| I
Db 9 KHYSSSRSGGGGGGGCG 27

RESULT 22

S65543
N:Contains: 3',5'-cyclic-nucleotide phosphodiesterase (EC 3.1.4.17), cAMP-specific, splice form II
C:Species: Drosophila melanogaster
C:Date: 28-Oct-1995 #sequence_revision 19-Jul-1996 #text_change 17-Nov-2000
C:Accession: S65543; S19662; S65542; S65544; A26651
R:Qiu, Y.; Chen, C.N.; Malone, T.; Richter, L.; Beckendorf, S.K.; Davis, R.L.
J. Mol. Biol. 222, 553-565, 1991
A:Title: Characterization of the memory gene dunce of Drosophila melanogaster.
A:Reference number: S19662; MUID:92085274
A:Accession: S65543
A:Molecule type: DNA
A:Residues: 1-777 <QIU>
A:Cross-references: EMBL:X55174
A:Accession: S19662
A:Molecule type: DNA
A:Residues: 137-777 <QI2>
A:Cross-references: EMBL:X55174
A:Accession: S65542
A:Molecule type: DNA
A:Residues: 'MOAQO', 86-87, 'IG', 90-91, 'QKYSRSLKNNRHTLANVR', 94-777 <QI3>
A:Cross-references: EMBL:X55174
A:Accession: S65544
A:Molecule type: DNA
A:Residues: 'MVCSCCCYNNRN', 4, 'P', 6, 'S', 94-777 <QI4>
A:Cross-references: EMBL:X55174
R:Chen, C.N.; Denome, S.; Davis, R.L.
Proc. Natl. Acad. Sci. U.S.A. 83, 9313-9317, 1986
A:Title: Molecular analysis of cDNA clones and the corresponding genomic coding sequence
A:Reference number: A26651; MUID:87092243
A:Accession: A26651
A:Molecule type: DNA
A:Residues: 416-777 <CHE>
A:Cross-references: GB:M14982; NID:g157278; PIDN:AAC34201.1; PID:g157280
C:Genetics:
A:Gene: FlyBase:dnc; dunce
A:Cross-references: FlyBase:FBgn0000479
A:Introns: 93/2; 125/3; 152/2; 200/2; 262/3; 294/1; 407/3; 496/2; 534/2; 588/3; 7
C:Superfamily: 3',5'-cyclic-nucleotide phosphodiesterase IB, calmodulin-dependent; 3',5'
C:Keywords: alternative splicing; phosphoric diester hydrolase
F:137-777/Product: cAMP-dependent 3',5'-cyclic-nucleotide phosphodiesterase, splice form
F:439-667/Domain: 3',5'-cyclic-nucleotide phosphodiesterase homology <CNP>

Query Match 31.4%; Score 61; DB 2; Length 777;
Best Local Similarity 75.0%; Pred. No. 35;
Matches 12; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGIEGP 26
: : : : : ||||| I
Db 748 ALRAGGGGGGGGMAP 763

RESULT 23

MWAXIC
myosin heavy chain IC - Acanthamoeba castellanii
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Acanthamoeba castellanii
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
C:Accession: A33891; C34448; A24146
R:Jung, G.; Korn, E.D.; Hammer III, J.A.
Proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987
A:Title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like and non-my
A:Reference number: A33891; MUID:88016163
A:Accession: A33891
A:Molecule type: DNA
A:Residues: 1-1168 <JUN>
A:Cross-references: GB:J02974; NID:g155624; PIDN:AAA27707.1; PID:g155625

A:Note: this gene and protein are called MIB in this paper
R:Brzeska, H.; Lynch, T.J.; Martin, B.; Korn, E.D.
J. Biol. Chem. 264, 19340-19348, 1989
A:Title: The localization and sequence of the phosphorylation sites of Acanthamoeba m
A:Reference number: A34448; MUID:90037074
A:Accession: C34448
A:Molecule type: protein
A:Residues: 308-314, 'X', 316-329 <BRZ>
C:Comment: In this protein, the coiled-coil rod-like region found in many myosin heav
he protein is globular and does not self-associate into filaments.
C:Genetics:
A:Gene: MIC
A:Introns: 1/3; 37/3; 60/2; 100/2; 153/3; 179/3; 208/2; 242/3; 321/3; 371/3; 4
C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 hom
C:Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosphoprotein
F:10-533/Domain: myosin motor domain homology <MMOT>
F:101-108/Region: nucleotide-binding motif A (P-loop)
F:543-564/Region: actin binding #status predicted
F:671-1168/Domain: carboxyl-terminal <CTD>
F:675-883/Region: basic
F:923-978/Region: alanine/glycine/proline-rich
F:983-1030/Domain: SH3 homology <SH3>
F:1034-1168/Region: alanine/glycine/proline-rich
F:107/Binding site: ATP (lys) #status predicted
F:311/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 31.4%; Score 61; DB 1; Length 1168;
Best Local Similarity 60.0%; Pred. No. 50;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 8 QWLAARAGGGGGIEGPT 27
: : : : : ||||| I
Db 920 QILCAKGGGGGGRGGPS 939

RESULT 24

A42687
neurotrophin-4 precursor - human
N:Alternate names: neurotrophin-5
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
C:Accession: A42687; JH0503
R:Ip, N.Y.; Ibanez, C.F.; Nye, S.H.; McClain, J.; Jones, P.F.; Gies, D.R.; Belluscio,
Proc. Natl. Acad. Sci. U.S.A. 89, 3060-3064, 1992
A:Title: Mammalian neurotrophin-4: structure, chromosomal localization, tissue distri
A:Reference number: A42687; MUID:92212967
A:Accession: A42687
A:Molecule type: DNA
A:Residues: 1-210 <IP1>
A:Cross-references: NID:g190264; PIDN:AAA60154.1; PID:g190265
A:Note: sequence extracted from NCBI backbone (NCBIN:93810, NCBIP:93811)
R:Berkemeier, L.R.; Winslow, J.W.; Kaplan, D.R.; Nikolics, K.; Goeddel, D.V.; Rosenth
Neuron 7, 857-866, 1991
A:Title: Neurotrophin-5: a novel neurotrophic factor that activates trk and trkB.
A:Reference number: JH0503; MUID:92075279
A:Accession: JH0503
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-210 <BER>
C:Comment: The neurotrophins stimulate autophosphorylation and transduce signals thro
C:Genetics:
A:Gene: GDB:NTF5
A:Cross-references: GDB:I34723; OMIM:162662
A:Map position: 19pter-19qter
C:Superfamily: nerve growth factor beta chain
C:Keywords: glycoprotein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-80/Domain: propeptide #status predicted <PRO>
F:81-210/Product: neurotrophin-4 #status predicted <NEU>
F:76/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 31.2%; Score 60.5; DB 2; Length 210;
Best Local Similarity 35.0%; Pred. No. 12;

Matches 14; Conservative 3; Mismatches 14; Indels 9; Gaps 1;
 QY 3 GTTLRWL-----AARAGGGGGGIEGPTLRQWLA 33
 Db 129 GSPLRQYFFETRCKADNAEEGPGAGGGCGVDRRHWS 168
 RESULT 25
 A48266
 protein-tyrosine kinase (EC 2.7.1.112) ltk - human
 N:Alternate names: protein-tyrosine kinase tyk1
 C:Species: Homo sapiens (man)
 C>Date: 16-Feb-1994 #sequence_revision 12-Apr-1996 #text_change 04-Feb-2000
 C:Accession: A48266; S17452; A60189
 R:Toyoshima, H.; Kozutsumi, H.; Maru, Y.; Hagiwara, K.; Furuya, A.; Mich, H.; Hanai, N.;
 Proc. Natl. Acad. Sci. U.S.A. 90, 5404-5408, 1993
 A:Title: Differently spliced cDNAs of human leukocyte tyrosine kinase receptor tyrosine
 A:Reference number: A48266; MUID:93296146
 A:Accession: A48266
 A:Molecule type: mRNA
 A:Residues: 1-864 <TOY>
 A:Cross-references: GB:D16105; NID:g440854; PIDN:BAA03679.1; PID:d1004194; PID:g440855
 A:Experimental source: placenta
 A:Note: sequence modified after extraction from NCBI backbone
 A:Note: the authors translated the codon CAG for residue 42 as Arg
 A:Note: sequence extracted from NCBI backbone (NCBIN:133811)
 R:Krolewski, J.J.; Dalla-Favera, R.
 EMBO J. 10, 2911-2919, 1991
 A:Title: The ltk gene encodes a novel receptor-type protein tyrosine kinase.
 A:Reference number: S17452; MUID:92007735
 A:Accession: S17452
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-41, R' 43-219, 'L' 221-273, 335-864 <KRO>
 A:Cross-references: EMBL:X60702; NID:g34419; PIDN:CAA43113.1; PID:g34420
 R:Krolewski, J.J.; Lee, R.; Eddy, R.; Shows, T.B.; Dalla-Favera, R.
 Oncogene 5, 277-282, 1990
 A:Title: Identification and chromosomal mapping of new human tyrosine kinase genes.
 A:Reference number: A60189; MUID:90191712
 A:Accession: A60189
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 608-716 <KR2>
 C:Genetics:
 A:Gene: GDB:LTK
 A:Cross-references: GDB:127768; OMIM:151520
 A:Map position: 15q15.1-15q15.2
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
 C:Keywords: alternative splicing; ATP; phosphotransferase; tyrosine-specific protein kin
 F:508-784/Domain: protein kinase homology <KIN>
 F:516-524/Region: protein kinase ATP-binding motif
 Query Match 31.2%; Score 60.5; DB 2; Length 864;
 Best Local Similarity 63.6%; Pred. No. 43;
 Matches 14; Conservative 1; Mismatches 2; Indels 5; Gaps 2;
 QY 2 EG-PTLRQWLAARAGGGGGG 22
 Db 196 EGVPGSRW----AGGGGGGG 213
 RESULT 26
 S69312
 probable membrane protein YLR338w - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein L8300.13-a
 C:Species: Saccharomyces cerevisiae
 C>Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 05-Nov-1999
 C:Accession: S69312
 R:Du, Z.
 submitted to the EMBL Data Library, January 1994
 A:Description: The sequence of S. cerevisiae cosmid 8300.
 A:Reference number: S69312
 A:Accession: S69312

A:Molecule type: DNA
 A:Residues: 1-285 <DUZ>
 A:Cross-references: EMBL:U19028; NID:g609380; PID:g2340034; GSPDB:GN00012; MIPS:YLR33
 C:Genetics:
 A:Gene: MIPS:YLR338w
 A:Map position: 12R
 C:Keywords: transmembrane protein
 F:142-158/Domain: transmembrane #status predicted <TM1>
 F:201-217/Domain: transmembrane #status predicted <TM2>
 Query Match 30.9%; Score 60; DB 2; Length 285;
 Best Local Similarity 57.9%; Pred. No. 18;
 Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 10 LAARAGGGGGGIEGPTL 28
 Db 236 LPPNAGGGGGGAGAPAI 254
 RESULT 27
 S20099
 transforming protein junD - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
 C:Accession: S20099
 R:Hartl, M.; Hutchins, J.T.; Vogt, P.K.
 Oncogene 6, 1623-1631, 1991
 A:Title: The chicken junD gene and its product.
 A:Reference number: S20099; MUID:92019832
 A:Accession: S20099
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-323 <HAR>
 A:Cross-references: EMBL:X60063; NID:g62927; PIDN:CAA42665.1; PID:g62928
 C:Superfamily: jun transforming protein; fos/jun DNA-binding domain homology
 C:Keywords: DNA binding; nucleus; transcription regulation
 F:237-277/Domain: fos/jun DNA-binding domain homology <FJD>
 Query Match 30.9%; Score 60; DB 2; Length 323;
 Best Local Similarity 72.2%; Pred. No. 20;
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 11 AARAGGGGGGIEGPTL 28
 Db 151 AAAAGGGGGGGGGEL 168
 RESULT 28
 KRMSEL
 keratin, 59K type I cytoskeletal - mouse
 N:Alternate names: 59-kda type I keratin
 C:Species: Mus musculus (house mouse)
 C>Date: 15-Nov-1984 #sequence_revision 04-Dec-1986 #text_change 10-Dec-1999
 C:Accession: A02940
 R:Krieg, T.M.; Schaefer, M.P.; Cheng, C.K.; Filpula, D.; Flaherty, P.; Steinert, P.M.;
 J. Biol. Chem. 260, 5867-5870, 1985
 A:Title: Organization of a type I keratin gene. Evidence for evolution of intermediat
 A:Reference number: A02940; MUID:85207552
 A:Accession: A02940
 A:Molecule type: DNA
 A:Residues: 1-569 <KRI>
 A:Cross-references: GB:L00193; GB:K00391; NID:g198625; PIDN:AAA39391.1; PID:g387397
 A:Note: Initiator Met not shown
 C:Comment: The authors translated the codon GAG for residue 41 as Gly
 A:Note: Fourier analysis has identified a 7-residue repeating pattern (heptad) bet
 forms a stable alpha-helical coiled coil but is interrupted by three short regions wi
 C:Comment: Most of the introns of the gene encoding this protein are located within t
 he sequence at or near the beginning of heptad repeats. Several of these sites are co
 C:Comment: The amino and carboxyl ends are rich in glycine, serine, and aromatic resi
 C:Genetics:
 A:Introns: 206/3; 234/2; 286/3; 340/3; 382/3; 455/2; 568/2
 C:Superfamily: cytoskeletal keratin
 C:Keywords: coiled coil; intermediate filament

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29977 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838C-26

Perfect score: 194

Sequence: 1 IEPTLRQWLARAGGGGGGIEPTLRQWLAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63.5	32.7	619	LAC1_NEUCR	P06811 neurospora
2	63.5	32.7	619	LAC2_NEUCR	P10574 neurospora
3	63	32.5	201	YR21_TRSVR	P25245 tomato ring
4	62.5	32.2	209	NT5_RAT	P34131 rattus norv
5	62.5	32.2	266	SCO2_HUMAN	O43819 homo sapien
6	61.5	31.7	969	PAC4_HUMAN	P29122 homo sapien
7	61	31.4	394	FXD3_CHICK	P79772 gallus gall
8	61	31.4	443	OC3N_HUMAN	P20265 homo sapien
9	61	31.4	445	OC3N_MOUSE	P31360 mus musculu
10	61	31.4	584	CNAL_DROME	P12252 drosophila
11	61	31.4	593	K1CJ_HUMAN	P13645 homo sapien
12	61	31.4	1168	MYSC_ACACA	P10569 acanthamoeb
13	61	31.4	1178	PHYB_SORBI	P93527 sorghum bic
14	60.5	31.2	210	NT5_HUMAN	P34130 homo sapien
15	60.5	31.2	864	KLTK_HUMAN	P29376 homo sapien
16	60	30.9	323	JUND_CHICK	P27921 gallus gall
17	60	30.9	348	SXL_CERCA	O61374 ceratitid c
18	60	30.9	440	DCO_DROME	O76324 drosophila
19	60	30.9	497	FXD2_HUMAN	O60548 homo sapien
20	60	30.9	569	K1CJ_MOUSE	P02535 mus musculu
21	60	30.9	888	KLTK_MOUSE	P08923 mus musculu
22	60	30.9	1322	SUS_DROME	P22293 drosophila
23	59.5	30.7	391	SOX1_MOUSE	P53783 mus musculu
24	59	30.4	367	BET3_MESAU	O09029 mesocricetu
25	59	30.4	401	HB9_HUMAN	P50219 homo sapien
26	59	30.4	485	ONC2_HUMAN	O95948 homo sapien
27	59	30.4	753	ZIN_HUMAN	O9nr13 homo sapien
28	59	30.4	757	ECR_LUCCU	O18531 lucilia cup
29	59	30.4	4499	DYHA_CHLRE	Q39610 chlamydomon
30	58.5	30.2	342	HXD9_HUMAN	P28356 homo sapien
31	58	29.9	339	HXD9_MOUSE	P28357 mus musculu
32	58	29.9	445	H3R_HUMAN	O9y5n1 homo sapien
33	58	29.9	476	EVX2_HUMAN	Q03828 homo sapien

ALIGNMENTS

RESULT 1

ID	LAC1_NEUCR	STANDARD;	PRT;	619 AA.
AC	P06811;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Laccase precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)			
DE	(Urishiol oxidase) (Laccase allele OR).			
GN	LACC.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Sordariales; Sordariaceae; Neurospora.			
OX	NCBI_TaxID=5141;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=88087214; PubMed=2961749;			
RA	Germann U.A., Mueller G., Hunziker P.E., Lerch K.;			
RT	"Characterization of two allelic forms of Neurospora crassa laccase.			
RT	Amino- and carboxyl-terminal processing of a precursor.";			
RL	J. Biol. Chem. 263:885-896(1988).			
RN	[2]			
RP	SEQUENCE OF 379-619 FROM N.A.			
RX	MEDLINE=87067412; PubMed=2947240;			
RA	Germann U.A., Lerch K.;			
RT	"Isolation and partial nucleotide sequence of the laccase gene from			
RT	Neurospora crassa: amino acid sequence homology of the protein to			
RT	human ceruloplasmin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).			
CC	-!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED			
CC	PRODUCTS (PROBABLE).			
CC	-!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2			
CC	H(2)O.			
CC	-!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU			
CC	CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE			
CC	3 OR COUPLED BINUCLEAR (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: Secreted (Potential).			
CC	-!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.			
CC	-!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M14554; AAA33590.1; .			
DR	EMBL; M18333; AAA33591.1; .			
DR	PIR; A28523; KSNCLD.			
DR	PIR; A29762; A29762.			
DR	InterPro; IPR001117; Cu-oxidase.			
DR	InterPro; IPR002355; MultiCu_oxidase2.			
DR	Pfam; PF00394; Cu-oxidase; 3.			
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.			

P31361 mus musculu
Q63262 rattus norv
P20264 homo sapien
O83933 treponema p
O54839 mus musculu
O55165 rattus norv
P25764 oryza sativ
P01317 bos taurus
P01318 ovis aries
P97273 cavia porce
P04574 sus scrofa
P04632 homo sapien

34 58 29.9 495 1 BRNL_MOUSE
35 58 29.9 497 1 BRNL_RAT
36 58 29.9 500 1 BRNL_HUMAN
37 58 29.9 517 1 Y967_TREPA
38 58 29.9 688 1 EOMD_MOUSE
39 58 29.9 796 1 KF3C_RAT
40 58 29.9 1171 1 PHYB_ORYSA
41 57.5 29.6 105 1 INS_BOVIN
42 57.5 29.6 105 1 INS_SHEEP
43 57 29.4 112 1 TTF1_CAVPO
44 57 29.4 266 1 CANS_PIG
45 57 29.4 268 1 CANS_HUMAN

DR PROSITE: PS0080; MULTICOPPER_OXIDASE2; 1.
 KW Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 KW Glycoprotein; Repeat.
 FT SIGNAL 1 21
 FT PROPEP 22 49
 FT CHAIN 50 606
 FT PROPEP 607 619
 FT DOMAIN 84 207
 FT DOMAIN 216 373
 FT DOMAIN 431 566
 FT METAL 144 144
 FT METAL 146 146
 FT METAL 189 189
 FT METAL 191 191
 FT METAL 477 477
 FT METAL 480 480
 FT METAL 482 482
 FT METAL 549 549
 FT METAL 550 550
 FT METAL 554 554
 FT METAL 559 559
 FT METAL 139 139
 FT CARBOHYD 282 282
 FT CARBOHYD 295 295
 FT CARBOHYD 340 340
 FT CARBOHYD 422 422
 FT CARBOHYD 444 444
 SQ SEQUENCE 619 AA; 68198 MW; FDEED6D78B65048E3 CRC64;
 Query Match 32.7%; Score 63.5; DB 1; Length 619;
 Best Local Similarity 57.7%; Pred. No. 9,9;
 Matches 15; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
 QY 7 ROWLAARAGGGGGGIEGPTLRQ-W 31
 || | ||||| || || |
 Db 39 RQDSQAERYGGGGGCGNSPTNRQW 64
 RESULT 2
 LAC2_NEUCR STANDARD; PRT; 619 AA.
 ID LAC2_NEUCR
 AC P10574;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Laccase precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
 DE {Urishiol oxidase} (Laccase allele TS).
 GN LACC.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88087214; PubMed=2961749;
 RA Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
 RT "Characterization of two allelic forms of Neurospora crassa laccase.
 RT Amino- and carboxyl-terminal processing of a precursor.";
 RL J. Biol. Chem. 263:885-896(1988).
 RC -|- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLY).
 CC -|- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzoquinone + 2
 CC H(2)O.
 CC -|- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -|- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -|- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -|- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M18334; AAA33592.1; -.
 DR PIR: B28523; KSNCLT.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; MultiCu_oxidase2.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
 DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
 KW Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 KW Glycoprotein; Repeat.
 FT SIGNAL 1 21
 FT PROPEP 22 49
 FT CHAIN 50 606
 FT PROPEP 607 619
 FT DOMAIN 84 207
 FT DOMAIN 216 373
 FT DOMAIN 431 566
 FT METAL 144 144
 FT METAL 146 146
 FT METAL 189 189
 FT METAL 191 191
 FT METAL 477 477
 FT METAL 480 480
 FT METAL 482 482
 FT METAL 549 549
 FT METAL 550 550
 FT METAL 554 554
 FT METAL 559 559
 FT METAL 139 139
 FT CARBOHYD 282 282
 FT CARBOHYD 295 295
 FT CARBOHYD 340 340
 FT CARBOHYD 422 422
 FT CARBOHYD 444 444
 SQ SEQUENCE 619 AA; 68120 MW; 0BB6CCDE18841145 CRC64;
 Query Match 32.7%; Score 63.5; DB 1; Length 619;
 Best Local Similarity 57.7%; Pred. No. 9,9;
 Matches 15; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
 QY 7 ROWLAARAGGGGGGIEGPTLRQ-W 31
 || | ||||| || || |
 Db 39 RQDSQAERYGGGGGCGNSPTNRQW 64
 RESULT 3
 YR21_TRSVR STANDARD; PRT; 201 AA.
 ID YR21_TRSVR
 AC P25245;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 20.2 kDa protein in RNA2.
 OS Tomato ringspot virus (isolate raspberry) (Tomrsv).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
 OC Nepovirus.
 OX NCBI_TaxID=12281;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91311402; PubMed=1856689;
 RA Rott M.E., Tremaine J.H., Rochon D.M.;
 RT "Nucleotide sequence of tomato ringspot virus RNA-2.";
 RL J. Gen. Virol. 72:1505-1514(1991).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: D12477; BAA02044.1; -
DR PIR: JQ1094; JQ1094.
DR HSP: P04002; LWFA.
KW Hypothetical protein.
FT DOMAIN 15 22 POLY-GLY.
FT DOMAIN 61 66 POLY-GLY.
FT DOMAIN 144 148 POLY-GLY.
SQ SEQUENCE 201 AA; 20194 MW; 9038506E18D7B450 CRC64;

Query Match 32.5%; Score 63; DB 1; Length 201;
Best Local Similarity 61.5%; Pred. No. 4.1;
Matches 16; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 13 RAGGGGGGGGIE-----GPTLRWLAA 34
DB 13 RAGGGGGGGGKEVFAGRTLLKVLKA 38

RESULT 4

NTS_RAT ID NTS_RAT STANDARD; PRT; 209 AA.
AC P34131;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurotrophin-5 precursor (NT-5) (Neurotrophic factor 5) (Neurotrophin-4)
DE (NT-4) (Neurotrophic factor 4).
GN NTF5 OR NTF4 OR NTF4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=92212967; PubMed=1313578;
RA IP N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
RA Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
RA Yancopoulos G.D.;
RA "Mammalian neurotrophin-4: structure, chromosomal localization,
RT tissue distribution, and receptor specificity";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=92075279; PubMed=1742028;
RA Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolic K., Goeddel D.V.,
RA Rosenthal A.;
RA "Neurotrophin-5: a novel neurotrophic factor that activates trk and
RT trkB";
RL Neuron 7:857-866(1991).
CC -!- FUNCTION: COULD SERVE AS A TARGET-DERIVED TROPHIC FACTOR FOR
CC SENSORY AND SYMPATHETIC NEURONS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, MUSCLE, OVARY, BRAIN,
CC HEART, STOMACH AND KIDNEY. EXPRESSED IN BOTH EMBRYO AND ADULT
CC TISSUES.
CC -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M86742; AAA41728.1; -
DR EMBL: S69323; AAB20548.1; -
DR PIR: JH0504; JH0504.
DR PIR: B42687; B42687.

DR HSP: P34130; 1BBM.
DR InterPro: IPR002072; NGF.
DR Pfam: PF00243; NGF; 1.
DR PRINTS: PR00268; NGF.
DR PRODOM: PD002052; NGF; 1.
DR SMART: SM00140; NGF; 1.
DR PROSITE: PS00248; NGF_1; 1.
DR PROSITE: PS02070; NGF_2; 1.
KW Growth factor; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 79
FT CHAIN 80 209 NEUROTROPHIN-5.
FT DISULFID 96 169 BY SIMILARITY.
FT DISULFID 140 198 BY SIMILARITY.
FT DISULFID 157 200 BY SIMILARITY.
FT CARBOHYD 75 75 N-LINKED (GLCNAC..) (POTENTIAL).
FT CONFLICT 177 177 R -> P (IN REF. 2).
SQ SEQUENCE 209 AA; 22332 MW; DF5112C05C5D5B85 CRC64;

Query Match 32.2%; Score 62.5; DB 1; Length 209;
Best Local Similarity 42.5%; Pred. No. 4.8;
Matches 17; Conservative 2; Mismatches 12; Indels 9; Gaps 2;

QY 3 GPTLRQWL-----AARAGGGG---GGGIEGPTLRWLAA 33
DB 128 GSPLRQYFFETRCRAESAGGPGVGCGRGVDRRHWS 167

RESULT 5

SC02_HUMAN ID SC02_HUMAN STANDARD; PRT; 266 AA.
AC 043819; O9UK87;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SC02 protein homolog, mitochondrial precursor.
GN SC02.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Monocytes.
RA Snink L.J., Burton J.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A., AND VARIANTS FIC LYS-140 AND PHE-225.
RX MEDLINE=20014747; PubMed=10545952;
RA Papadopoulos L.C., Sue C.M., Davidson M.M., Tanji K., Nishino I.,
RA Sadlock J.E., Krishna S., Walker W., Selby J., Gerum D.M.,
RA Van Coster R., Lyon G., Scalsis E., Lebel R., Kaplan P., Shanske S.,
RA De Vivo D.C., Bonilla E., Hirano M., DiMauro S., Schon E.A.;
RA "Fatal infantile cardioencephalomyopathy with COX deficiency and
RT mutations in SC02, a COX assembly gene";
RL Nat. Genet. 23:333-337(1999).
CC -!- FUNCTION: THOUGHT TO PLAY A ROLE IN EITHER MITOCHONDRIAL COPPER
CC TRANSPORT OR INSERTION OF COPPER INTO THE ACTIVE SITE OF COX.
CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- DISEASE: DEFECTS IN SC02 ARE THE CAUSE OF FATAL INFANTILE
CC CARDIOENCEPHALOMYOPATHY WITH COX DEFICIENCY. THIS DISEASE IS
CC CHARACTERIZED BY HYPERTROPHIC CARDIOMYOPATHY, LACTIC ACIDOSIS, AND
CC GLIOSIS. HEART AND SKELETAL MUSCLE SHOW REDUCTIONS IN COX
CC ACTIVITY, WHEREAS LIVER AND FIBROBLASTS SHOW MILD COX
CC DEFICIENCIES.
CC -!- SIMILARITY: BELONGS TO THE SC01/2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: AF177385; AAF05313.1; -
 DR EMBL: AL021683; CAA16671.1; -
 DR MIM: 604272; -
 DR MIM: 604377; -
 DR MIM: 220110; -
 DR InterPro: IPR003782; SCOL_Senc.
 DR Pfam: PF02630; SCOL_Senc; 1.
 KW Mitochondrion; Trans peptide; Disease mutation; Polymorphism.
 FT TRANSIT 1 41 MITOCHONDRION (POPELIAL).
 FT CHAIN 42 266 SC02 PROTEIN HOMOLOG.
 FT VARIANT 20 20 R -> P (IN DBSNP:140523).
 FT /FTID=VAR_011738.
 FT VARIANT 140 140 E -> K (IN FIC).
 FT /FTID=VAR_008874.
 FT VARIANT 225 225 S -> F (IN FIC).
 FT /FTID=VAR_008875.
 FT /FTID=VAR_008875.
 SQ SEQUENCE 266 AA; 29810 MW; BC2F40E057329BF3 CRC64;
 Query Match 32.2%; Score 62.5; DB 1; Length 266;
 Best Local Similarity 37.5%; Pred. No. 5.9;
 Matches 18; Conservative 2; Mismatches 11; Indels 17; Gaps 2;
 QY 6 LKWLAAARAGGG--GGGGGEGPTLR-----QWLAAAR 36
 DB 33 LRSWLLSRGPAETGGGQPGGGLTRLLTGLFGAGLGGAWLALRA 80
 RESULT 6
 PAC4_HUMAN STANDARD; PRT: 969 AA
 ID PAC4_HUMAN Q15100; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8; Q9UEJ9;
 AC P29122; Q15099; Q15100; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8; Q9UEJ9;
 AC Q9UEG7; Q944G9; Q944H0; Q944H1;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
 DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like proprotein
 DE convertase 4) (SPC4).
 DE PACE4.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).
 RC TISSUE=Hepatoma, and kidney;
 RX MEDLINE=92075167; PubMed=1741956;
 RA Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,
 RA Barr P.J.;
 RA "Identification of a second human subtilisin-like protease gene in
 RT the fes/fps region of chromosome 15";
 RL DNA Cell Biol. 10:757-769(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).
 RC TISSUE=Placenta;
 RX MEDLINE=94235049; PubMed=8179631;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RA "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms";
 RT PACE4 isoforms";
 RL Biochem. Biophys. Res. Commun. 200:943-950(1994).
 RN [3]
 RP ERATUM
 RX MEDLINE=95071480; PubMed=7980617;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RA "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms";
 RT PACE4 isoforms";
 RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM PACE4A-II).
 RC TISSUE=Placenta;
 RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,
 RA Matsuda Y.;
 RT "Identification of a novel PACE4 isoform, PACE4E";
 RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).
 RC TISSUE=Cerebellum;
 RX MEDLINE=97335942; PubMed=9192737;
 RA Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,
 RA Akamatsu T., Nagamune H., Matsuda Y.;
 RT "A novel human PACE4 isoform, PACE4E is an active processing protease
 RT containing a hydrophobic cluster at the carboxy terminus";
 RL J. Biochem. 121:941-948(1997).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; A-II; CS; D; E-I; E-II).
 RC MEDLINE=98021085; PubMed=9378725;
 RA Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
 RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
 RT "Genomic organization and alternative splicing of human PACE4 (SPC4),
 RT kexin-like processing endoprotease";
 RL J. Biochem. 122:438-452(1997).
 RN [7]
 RP ALTERNATIVE SPLICING (ISOFORM PACE4CS).
 RC MEDLINE=97064242; PubMed=8906861;
 RA Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
 RT "Functional analysis of human PACE4-A and PACE4-C isoforms:
 RT identification of a new PACE4-CS isoform";
 RL FEBS Lett. 396:31-36(1996).
 RN [8]
 RP CHARACTERIZATION.
 RC MEDLINE=99233559; PubMed=10215603;
 RA Sucic J.F., Moehring J.M., Innocencio N.M., Luchini J.W.,
 RA Moehring T.J.; PACE4 is Ca2+-dependent and temperature-sensitive and
 RT can partly rescue the phenotype of a furin-deficient cell strain";
 RL Biochem. J. 339:639-647(1999).
 RN [9]
 RP PROCESSING.
 RC MEDLINE=98408849; PubMed=9738469;
 RA Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
 RA Tsuji A., Matsuda Y.;
 RT "Biosynthetic processing and quaternary interactions of proprotein
 RT convertase SPC4 (PACE4)";
 RL FEBS Lett. 434:155-159(1998).
 CC -!- FUNCTION: LIKELY TO REPRESENT AN ENDOPROTEASE ACTIVITY WITHIN THE
 CC CONSTITUTIVE SECRETORY PATHWAY, WITH UNIQUE RESTRICTED
 CC DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
 CC AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS,
 CC WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -!- COFACTOR: PACE4A IS PROBABLY CALCIUM-DEPENDENT.
 CC -!- SUBUNIT: THE PACE4A-I PRECURSOR PROTEIN SEEMS TO EXIST IN THE
 CC RETICULUM ENDOPLASMIC AS BOTH A MONOMER AND A DIMER-SIZED COMPLEX
 CC WHEREAS MATURE PACE4A-I EXISTS ONLY AS A MONOMER, SUGGESTING THAT
 CC PROPEPTIDE CLEAVAGE AFFECTS ITS TERTIARY OR QUATERNARY STRUCTURE.
 CC -!- SUBCELLULAR LOCATION: PACE4A-I AND PACE4A-II ARE SECRETED. PACE4C
 CC AND PACE4CS ARE NOT SECRETED AND REMAIN PROBABLY IN ZYMOGEN FORM
 CC IN ENDOPLASMIC RETICULUM. PACE4E-I AND PACE4E-II ARE RETAINED
 CC INTRACELLULARLY PROBABLY THROUGH A HYDROPHOBIC CLUSTER IN THEIR C-
 CC TERMINUS. PACE4B MIGHT BE SECRETED.
 CC -!- ALTERNATIVE PRODUCTS: 8 ISOFORMS; PACE4A-I/PACE4 (SHOWN HERE),
 CC PACE4A-II, PACE4B/PACE4.1, PACE4C, PACE4CS, PACE4D, PACE4E-I AND
 CC PACE4E-II; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS PACE4B,
 CC C, CS AND D MIGHT BE ENZYMATICALLY INACTIVE.
 CC -!- TISSUE SPECIFICITY: EACH PACE4 ISOFORM EXHIBITS A UNIQUE
 CC RESTRICTED DISTRIBUTION. PACE4A-I IS EXPRESSED IN HEART, BRAIN,
 CC PLACENTA, LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS, BUT AT
 CC COMPARATIVELY HIGHER LEVELS IN THE LIVER. PACE4A-II IS AT LEAST
 CC COMPRESSED IN PLACENTA. PACE4B WAS ONLY FOUND IN THE EMBRYONIC
 CC KIDNEY CELL LINE FROM WHICH IT WAS ISOLATED. PACE4C AND PACE4D ARE

EXPRESSED IN PLACENTA. PAC4E-I IS EXPRESSED IN CEREBELLUM,
PLACENTA AND PITUITARY. PAC4E-II IS AT LEAST PRESENT IN
CEREBELLUM.
-!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM. ISOFORM PAC24D LACKS THE PROPEPTIDE DOMAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M80482; AAA59998.1; -
CC EMBL; AB001914; BAA21620.1; -
CC EMBL; AB001898; BAA21620.1; JOINED.
CC EMBL; AB001900; BAA21620.1; JOINED.
CC EMBL; AB001901; BAA21620.1; JOINED.
CC EMBL; AB001902; BAA21620.1; JOINED.
CC EMBL; AB001903; BAA21620.1; JOINED.
CC EMBL; AB001904; BAA21620.1; JOINED.
CC EMBL; AB001905; BAA21620.1; JOINED.
CC EMBL; AB001914; BAA21621.1; -
CC EMBL; AB001898; BAA21621.1; JOINED.
CC EMBL; AB001900; BAA21621.1; JOINED.
CC EMBL; AB001901; BAA21621.1; JOINED.
CC EMBL; AB001902; BAA21621.1; JOINED.
CC EMBL; AB001903; BAA21621.1; JOINED.
CC EMBL; AB001904; BAA21621.1; JOINED.
CC EMBL; AB001905; BAA21621.1; JOINED.
CC EMBL; AB001906; BAA21621.1; JOINED.
CC EMBL; AB001907; BAA21621.1; JOINED.
CC EMBL; AB001908; BAA21621.1; JOINED.
CC EMBL; AB001909; BAA21621.1; JOINED.
CC EMBL; AB001914; BAA21622.1; -
CC EMBL; AB001901; BAA21622.1; JOINED.
CC EMBL; AB001902; BAA21622.1; JOINED.
CC EMBL; AB001903; BAA21622.1; JOINED.
CC EMBL; AB001904; BAA21622.1; JOINED.
CC EMBL; AB001905; BAA21622.1; JOINED.
CC EMBL; AB001906; BAA21622.1; JOINED.
CC EMBL; AB001907; BAA21622.1; JOINED.
CC EMBL; AB001908; BAA21622.1; JOINED.
CC EMBL; AB001914; BAA21623.1; -
CC EMBL; AB001898; BAA21623.1; JOINED.
CC EMBL; AB001900; BAA21623.1; JOINED.
CC EMBL; AB001901; BAA21623.1; JOINED.
CC EMBL; AB001902; BAA21623.1; JOINED.
CC EMBL; AB001903; BAA21623.1; JOINED.
CC EMBL; AB001904; BAA21623.1; JOINED.
CC EMBL; AB001905; BAA21623.1; JOINED.
CC EMBL; AB001906; BAA21623.1; JOINED.
CC EMBL; AB001907; BAA21623.1; JOINED.
CC EMBL; AB001908; BAA21623.1; JOINED.
CC EMBL; AB001909; BAA21623.1; JOINED.
CC EMBL; AB001914; BAA21624.1; -
CC EMBL; AB001898; BAA21624.1; JOINED.
CC EMBL; AB001900; BAA21624.1; JOINED.
CC EMBL; AB001901; BAA21624.1; JOINED.
CC EMBL; AB001902; BAA21624.1; JOINED.
CC EMBL; AB001903; BAA21624.1; JOINED.
CC EMBL; AB001904; BAA21624.1; JOINED.
CC EMBL; AB001905; BAA21624.1; JOINED.
CC EMBL; AB001906; BAA21624.1; JOINED.
CC EMBL; AB001907; BAA21624.1; JOINED.
CC EMBL; AB001908; BAA21624.1; JOINED.
CC EMBL; AB001910; BAA21624.1; JOINED.
CC EMBL; AB001911; BAA21624.1; JOINED.

DR EMBL; AB001912; BAA21624.1; JOINED.
DR EMBL; AB001913; BAA21624.1; JOINED.
DR EMBL; AB001914; BAA21625.1; -
DR EMBL; AB001898; BAA21625.1; JOINED.
DR EMBL; AB001900; BAA21625.1; JOINED.
DR EMBL; AB001901; BAA21625.1; JOINED.
DR EMBL; AB001902; BAA21625.1; JOINED.
DR EMBL; AB001903; BAA21625.1; JOINED.
DR EMBL; AB001904; BAA21625.1; JOINED.
DR EMBL; AB001905; BAA21625.1; JOINED.
DR EMBL; AB001906; BAA21625.1; JOINED.
DR EMBL; AB001907; BAA21625.1; JOINED.
DR EMBL; AB001908; BAA21625.1; JOINED.

Query Match 31.7%; Score 61.5; DB 1; Length 969;
Best Local Similarity 48.3%; Pred. No. 23;
Matches 14; Conservative 1; Mismatches 7; Indels 7; Gaps 1;

Qy 11 AARAGGGGGGIEGPTLR-----QWL 32
||| ||| ||| ||| |||
Db 24 AAGAGGAGGAGGAGGPGFRPLAPRPWRL 52

RESULT 7
FXD3_CHICK
ID FXD3_CHICK STANDARD; PRT; 394 AA.
AC P79772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein D3 (HNF3/FH transcription factor genesis) (Winged
DE helix protein CWH-3).
GN FOXD3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=97141794; PubMed=8988052;
RA Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;
RT "Aberrant cell growth induced by avian winged helix proteins.";
RL Cancer Res. 57:123-129(1997).
CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U37274; AAC60066.1; -
CC HSSP; Q63245; 2HFH.
DR TRANSFAC; T02495;
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 67 70 POLY-ALA.
FT DOMAIN 80 91 POLY-GLY.
FT DOMAIN 100 106 POLY-ALA.
FT DOMAIN 117 211 FORK-HEAD.
SQ SEQUENCE 394 AA; 40995 MW; 324A4B36B9E31899 CRC64;

Query Match 31.4%; Score 61; DB 1; Length 394;
 Best Local Similarity 84.6%; Pred No. 12;
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 RAGGGGGGGGIEG 25
 | | | | | | | | | |
 Db 82 RGGGGGGGGGEG 94

RESULT 8
 OC3N_HUMAN STANDARD; PRT; 443 AA.
 AC P20265; Q14960;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nervous-system specific octamer-binding transcription factor N-OCT 3
 DE (Brain-specific homeobox/POU domain protein 2) (BRN-2 protein)
 DE [Contains: N-OCT 5A; N-OCT 5B].
 GN POU3F2 OR BRN2 OR OTF7 OR OCT7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93181199; PubMed=8441633;
 RA Schreiber E., Tobler A., Malipiero U., Schaffner W., Fontana A.;
 RT "cDNA cloning of human N-Oct3, a nervous-system specific POU domain
 RT transcription factor binding to the octamer DNA motif.";
 RL Nucleic Acids Res. 21:253-258(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95380176; PubMed=7651733;
 RA Angus J., Thomson F., Murphy K., Baker E., Sutherland G.R.,
 RA Parsons P.G., Sturm R.A.;
 RT "The brn-2 gene regulates the melanocytic phenotype and tumorigenic
 RT potential of human melanoma cells.";
 RL Oncogene 11:691-700(1995).
 RN [3]
 RP SEQUENCE OF 280-404 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=89295573; PubMed=2739723;
 RA He X., Treacy M.N., Simmons D.M., Ingraham H.A., Swanson L.W.,
 RA Rosenfeld M.G.;
 RT "Expression of a large family of POU-domain regulatory genes in
 RT mammalian brain development.";
 RL Nature 340:35-42(1989).
 CC [1] FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
 CC RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,
 CC ('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION
 CC OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER
 CC THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II
 CC PROMOTERS (BY SIMILARITY).
 CC [1] SUBCELLULAR LOCATION: Nuclear.
 CC [1] ALTERNATIVE PRODUCTS: 3 ISOFORMS: N-OCT 3 (SHOWN HERE), N-OCT 5A
 CC AND N-OCT 5B; ARE PRODUCED BY ALTERNATIVE INITIATION
 CC [1] TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL
 CC CELL LINEAGE.
 CC [1] SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
 CC TO CLASS-3 POU.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z11933; CAA77990.1; -.
 DR EMBL; L37868; AAB59611.1; -.
 DR PIR; S05043; S05043.
 DR PIR; S29334; S29334.
 DR HSP; P14859; I0CT.
 DR TRANSFAC; T00630; -.
 DR MIM; 600494; -.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR000327; POU.
 DR Pfam; PF00046; homeobox; 1.
 DR Pfam; PF00157; pou; 1.
 DR PRINTS; PD00028; POU DOMAIN.
 DR PRODOM; PD000583; POU; 1.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00352; POU; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS00035; POU_1; 1.
 DR PROSITE; PS00465; POU_2; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 KW DNA-binding; Nuclear protein; Homeobox; Transcription regulation;
 KW Activator; Alternative initiation.
 FT CHAIN 1 443 N-OCT 3.
 FT CHAIN 181 443 N-OCT 5A.
 FT CHAIN 200 443 N-OCT 5B.
 FT INIT_MET 181 181 FOR N-OCT 5A.
 FT INIT_MET 200 200 FOR N-OCT 5B.
 FT DOMAIN 68 90 POLY-GLY.
 FT DOMAIN 125 149 POLY-GLN.
 FT DOMAIN 266 336 POU.
 FT DNA_BIND 354 413 HOMEBOX.
 FT CONFLICT 26 26 A -> G (IN REF. 2).
 SQ SEQUENCE 443 AA; 46921 MW; 2CAC852328334A66 CRC64;

Query Match 31.4%; Score 61; DB 1; Length 443;
 Best Local Similarity 66.7%; Pred. No. 13;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 QWLAARAGGGGGGG 22
 | | | | | | | | | |
 Db 60 QWITALSHGGGGGG 74

RESULT 9
 OC3N_MOUSE STANDARD; PRT; 445 AA.
 AC P31360;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Nervous-system specific octamer-binding transcription factor N-OCT 3
 DE (Brain-specific homeobox/POU domain protein 2) (BRN-2 protein).
 GN POU3F2 OR OTF7 OR BRN2 OR BRN-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92228768; PubMed=1565620;
 RA Hara Y., Rovescalli C., Kim Y., Nirenberg M.;
 RT "Structure and evolution of four POU domain genes expressed in mouse
 RT brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3280-3284(1992).
 CC [1] FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE
 CC RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,
 CC ('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION
 CC OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER
 CC THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II
 CC PROMOTERS (BY SIMILARITY).
 CC [1] SUBCELLULAR LOCATION: Nuclear.
 CC [1] TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL
 CC CELL LINEAGE.

RP [1] SEQUENCE FROM N.A.
RX MEDLINE-89125611; PubMed-2464696;
RA Rieger M., Franke W.W.;
RT "Identification of an orthologous mammalian cytokeratin gene. High
RT degree of intron sequence conservation during evolution of human
RT cytokeratin 10.";
RL J. Mol. Biol. 204:841-856(1988).
RN [2]
RP SEQUENCE OF 130-593 FROM N.A.
RX MEDLINE-88122104; PubMed-2448602;
RA Darmon M.Y., Semat A., Darmon M.C., Vasseur M.;
RT "Sequence of a cDNA encoding human keratin No 10 selected according
RT to structural homologies of keratins and their tissue-specific
RT expression.";
RL Mol. Biol. Rep. 12:277-283(1987).
RN [3]
RP SEQUENCE OF 197-593 FROM N.A.
RX MEDLINE-92339897; PubMed-1378806;
RA Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P.,
RA Kisselev L.L.;
RT "Exons I and VII of the gene (Ker10) encoding human keratin 10
RT undergo structural rearrangements within repeats.";
RL Gene 116:245-251(1992).
RN [4]
RP SEQUENCE OF 180-184 AND 577-589.
RX TISSUE-Keratinocytes;
RC MEDLINE-93162043; PubMed-1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
RN [5]
RP VARIANTS EHK HIS-156.
RX MEDLINE-922386600; PubMed-1381287;
RA Cheng J., Syder A.J., Yu Q.-C., Letai A., Paller A.S., Fuchs E.;
RT "The genetic basis of epidermolysis hyperkeratosis: a disorder of
RT differentiation-specific epidermal keratin genes.";
RL Cell 70:811-819(1992).
RN [6]
RP VARIANTS
RX MEDLINE-92141228; PubMed-1371013;
RA Korge B.P., Gan S.-Q., McBride O.W., Mischke D., Steinert P.M.;
RT "Extensive size polymorphism of the human keratin 10 chain resides in
RT the C-terminal V2 subdomain due to variable numbers and sizes of
RT glycine loops.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).
RN [7]
RP VARIANTS EHK HIS-156 AND SER-161.
RX MEDLINE-92376531; PubMed-1380725;
RA Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT "Mutations in the rod domains of keratins 1 and 10 in epidermolysis
RT hyperkeratosis.";
RL Science 257:1128-1130(1992).
RN [8]
RP VARIANTS EHK HIS-154; CYS-156; HIS-156; ASP-160 AND GLN-442.
RX MEDLINE-94136477; PubMed-7508181;
RA Chipev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L.,
RA Compton J.G., Bale S.J.;
RT "Preferential sites in keratin 10 that are mutated in epidermolysis
RT hyperkeratosis.";
RL Am. J. Hum. Genet. 54:179-190(1994).
RN [9]
RP VARIANTS EHK ARG-150; CYS-156 AND GLU-439, AND VARIANT SER-126.
RX MEDLINE-94216497; PubMed-7512983;
RA Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;
RT "Genetic mutations in the K1 and K10 genes of patients with
RT epidermolysis hyperkeratosis. Correlation between location and
RT disease severity.";
RL J. Clin. Invest. 93:1533-1542(1994).
RN [10]
RP VARIANTS EHK ASN-160.
RX MEDLINE-94117868; PubMed-7507150;
RA Rothnagel J.A., Longley M.A., Holder R.A., Kuster W., Roop D.R.;
RT "Prenatal diagnosis of epidermolysis hyperkeratosis by direct gene
RT sequencing.";
RL J. Invest. Dermatol. 102:13-16(1994).
RN [11]
RP VARIANTS EHK PRO-156 AND SER-156.
RX MEDLINE-94117870; PubMed-7507152;
RA McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,
RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA Morley S.M.;
RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT congenital ichthyosiform erythroderma (BCIE).";
RL J. Invest. Dermatol. 102:24-30(1994).
RN [12]
RP VARIANTS EHK THR-150.
RX MEDLINE-95059228; PubMed-7526210;
RA Paller A.S., Syder A.J., Chan Y.-M., Yu Q.-C., Hutton M.E., Tadini G.,
RA Fuchs E.;
RT "Genetic and clinical mosaicism in a type of epidermal nevus.";
RL New Engl. J. Med. 331:1408-1415(1994).
RN [13]
RP VARIANTS AEI THR-446.
RX MEDLINE-99072665; PubMed-9856845;
RA Suga Y., Duncan K.O., Heald P.W., Roop D.R.;
RT "A novel helix termination mutation in keratin 10 in annular
RT epidermolysis hyperkeratosis, a variant of bullous congenital
RT ichthyosiform erythroderma.";
RL J. Invest. Dermatol. 111:1220-1223(1998).
RN [14]
RP VARIANTS EHK SER-160.
RX MEDLINE-99215719; PubMed-10201536;
RA Arin M.J., Longley M.A., Anton-Lamprecht I., Kurze G., Huber M.,
RA Hohl D., Rothnagel J.A., Roop D.R.;
RT "A novel substitution in keratin 10 in epidermolysis hyperkeratosis.";
RL J. Invest. Dermatol. 112:506-508(1999).
RN [15]
RP SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
RT KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.
CC [1-] TISSUE SPECIFICITY: SEEN IN ALL SUPRABASAL CELL LAYERS INCLUDING
CC STRATUM CORNEUM.
CC [1-] POLYMORPHISM: A NUMBER OF ALLELES ARE KNOWN THAT MAINLY DIFFER IN
CC THE GLY-RICH REGION (POSITIONS 490-560).
CC [1-] DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF EPIDERMOLYTIC
CC HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC ICHTHYOSIFORM ERYTHRODERMA (BCIE)); A HEREDITARY SKIN DISORDER
CC CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC CORNEUM. AT BIRTH, AFFECTED INDIVIDUALS USUALLY PRESENT WITH
CC REDNESS, BLISTERS AND SUPERFICIAL EROSIONS DUE TO CYTOLYSIS.
CC WITHIN A FEW WEEKS, THE ERYTHRODERMA AND BLISTER FORMATION
CC DIMINISH AND HYPERKERATOSIS DEVELOP. TRANSMISSION IS AUTOSOMAL
CC DOMINANT, BUT MOST CASES ARE SPORADIC.
CC [1-] DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF ANNULAR EPIDERMOLYTIC
CC ICHTHYOSIS (AEI), A DISTINCT PHENOTYPIC VARIANT OF EPIDERMOLYTIC
CC HYPERKERATOSIS. IT RESEMBLES CLINICAL AND HISTOLOGIC FEATURES OF
CC BOTH EPIDERMOLYTIC HYPERKERATOSIS AND ICHTHYOSIS BULLOSA OF
CC STEMMEN.
CC [1-] MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II
CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].
CC [1-] SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC [1-] CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY IN
CC POSITIONS 513 TO 555.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC EMBL; X14487; CAA32649.1; -.

```

DR EMBL; M19156; AAA59468.1; -.
DR EMBL; M77663; AAA59199.1; -.
DR EMBL; L20218; AAB59438.1; -.
DR EMBL; L20219; AAB59439.1; -.
DR PIR; S02158; KRH00.
DR AarHus/Ghent-2DPAGE; 7405; IEF.
DR MIM; 148080; -.
DR MIM; 113800; -.
DR InterPro; IPR001664; IF.
DR InterPro; IPR002957; Keratin_I.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PRO1248; TYPEKERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin; Disease mutation;
KW Polymorphism.
FT DOMAIN 1 145 HEAD.
FT DOMAIN 146 456 ROD.
FT DOMAIN 457 593 TAIL.
FT DOMAIN 146 181 COIL 1A.
FT DOMAIN 182 202 LINKER 1.
FT DOMAIN 203 294 COIL 1B.
FT DOMAIN 295 317 LINKER 12.
FT DOMAIN 318 456 COIL 2.
FT DOMAIN 6 144 GLY/PHE/SER-RICH.
FT DOMAIN 451 590 GLY/SER-RICH.
FT VARIANT 126 126 G -> S.
FT VARIANT 150 150 M -> R (IN EHK).
FT VARIANT 150 150 M -> T (IN EHK).
FT VARIANT 154 154 N -> H (IN EHK).
FT VARIANT 156 156 R -> H (IN EHK).
FT VARIANT 156 156 R -> C (IN EHK).
FT VARIANT 156 156 R -> P (IN EHK).
FT VARIANT 156 156 R -> S (IN EHK).
FT VARIANT 160 160 Y -> D (IN EHK; SEVERE PHENOTYPE).
FT VARIANT 160 160 /FTId=VAR_003831.

Query Match 31.4%; Score 61; DB 1; Length 593;
Best Local Similarity 52.6%; Pred. No. 17;
Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 7 ROWLAARAGGGGGGGG 25
DB 9 KHYSSRSRGGGGGGGCGG 27

RESULT 12
MYSC-ACACA STANDARD; PRT; 1168 AA.
AC P10569;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IC heavy chain.
GN MIC.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88016163; PubMed=3477803;
RA Jung G., Korn E.D., Hammer J.A. III;
RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
and non-myosin-like sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724 (1987).
RN [2]

RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86259656; PubMed=3014500;
RA Hammer J.A. III, Jung G., Korn E.D.;
RT "Genetic evidence that Acanthamoeba myosin I is a true myosin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4655-4659 (1986).
RN [3]
RP PHOSPHORYLATION SITE.
RX MEDLINE=90037074; PubMed=2530230;
RA Brzeska H., Lynch T.J., Martin B., Korn E.D.;
RT "The localization and sequence of the phosphorylation sites of
Acanthamoeba myosins I. An improved method for locating the
phosphorylated amino acid.";
RL J. Biol. Chem. 264:19340-19348 (1989).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -!- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
MYOSIN I CAN THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND VICE
VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER WITH
THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE MOLECULES
OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
CC -!- MISCELLANEOUS: THIS ORGANISM EXPRESSES AT LEAST THREE ISOFORMS OF
MYOSIN I HEAVY-CHAIN, ENCODED BY GENES MIA, MIB, AND MIC.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE MYOSIN IB.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; J02974; AAA27707.1; -.
PIR; A33891; MWXIC.
DR HSSP; P08799; ILVK.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Myosin; ATP-binding; Phosphorylation; Multigene family; SH3 domain.
FT DOMAIN 1 670 MYOSIN HEAD-LIKE.
FT DOMAIN 671 922 TAIL HOMOLOGY REGION 1 (TH.1).
FT DOMAIN 923 975 GLY/PRO/ALA-RICH (TH.2).
FT DOMAIN 976 1035 SH3.
FT DOMAIN 1036 1168 GLY/PRO/ALA-RICH (TH.2).
FT NP_BIND 101 108 ATP (POTENTIAL).
FT MOD_RES 311 311 PHOSPHORYLATION.
SQ SEQUENCE 1168 AA; 127309 MW; D07084B373A37A32 CRC64;

Query Match 31.4%; Score 61; DB 1; Length 1168;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 8 OWLAARAGGGGGGGGIEGPT 27
DB 920 QILGAKGGGGGGGRGGPS 939

RESULT 13
PHYB-SORBI STANDARD; PRT; 1178 AA.
ID PHYB-SORBI
AC P93527;
DT 16-OCT-2001 (Rel. 40, Created)

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Phytochrome B.
 DN PHYB OR MA3.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Sorghum.
 ON NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. 58M;
 RX MEDLINE=20188796; PubMed=10723737;
 RA Alba R., Kelmenson P.M., Cordonnier-Pratt M.-M., Pratt L.H.;
 RT "The phytochrome gene family in tomato and the rapid differential
 evolution of this family in angiosperms";
 RL Mol. Biol. Evol. 17:362-373(2000).
 RN [2]
 RP SEQUENCE OF 208-1178 FROM N.A.
 RC STRAIN=CV. 58M;
 RX MEDLINE=97198556; PubMed=9046599;
 RA Childs K.L., Miller F.R., Cordonnier-Pratt M.-M., Pratt L.H.,
 RA Morgan P.W., Mullet J.E.;
 RT "The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a
 phytochrome B";
 RL Plant Physiol. 113:611-619(1997).
 CC -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
 ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
 MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
 ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
 PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
 RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
 RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
 GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
 BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
 PROTOCHLOROPHYLLIDE REDUCTASE, RNA, ETC. IT ALSO CONTROLS THE
 EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION (BY
 SUBUNIT).
 CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -!- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
 CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF182394; AAB41398.2; -
 DR InterPro: IPR003018; GAF.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR003661; His_kinA.
 DR InterPro: IPR000014; PAS.
 DR InterPro: IPR001294; Phytochrome.
 DR Pfam: PF01590; GAF; 1.
 DR Pfam: PF0518; HATPase_c; 1.
 DR Pfam: PF00989; PAS; 2.
 DR Pfam: PF00360; phytochrome; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PR01033; PHYTOCHROME.
 DR SMART: SM00065; GAF; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00388; HisKA; 1.
 DR SMART: SM00091; PAS; 2.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAS; 2.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS50046; PHYTOCHROME_2; 1.
 KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
 KW Repeat; Multigene family.
 FT DOMAIN 668 739 PAS 1.
 FT DOMAIN 802 873 PAS 2.
 FT DOMAIN 950 1170 HISTIDINE KINASE.
 FT DOMAIN 23 31 POLY-HIS.
 FT DOMAIN 43 54 POLY-GLY.
 FT BINDING 372 372 CHROMOPHORE (BY SIMILARITY).
 SQ SEQUENCE 1178 AA; 129136 MW; C406DF221197B93F CRC64;
 Query Match 31.4%; Score 61; DB 1; Length 1178;
 Best Local Similarity 75.0%; Pred. No. 31;
 Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 12 ARAGGGGGGGGIEGPT 27
 :|||||
 DB 40 SRAGGGGGGGGGGT 55
 RESULT 14
 NT5_HUMAN
 ID NT5_HUMAN STANDARD; PRT; 210 AA.
 AC P34130;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurotrophin-5 precursor (NT-5) (Neurotrophic factor 5) (Neurotrophin-4)
 DE (NT-4) (Neurotrophic factor 4).
 GN NT5 OR NTF4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eumalia; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX MEDLINE=92212967; PubMed=1313578;
 RA Ip N.Y., Ibanez C.F., Nye S.H., McClain J., Jones P.F., Gies D.R.,
 RA Belluscio L., le Beau M.M., Espinosa R. III, Squinto S.P., Persson H.,
 RA Yancopoulos G.D.;
 RT "Mammalian neurotrophin-4: structure, chromosomal localization,
 tissue distribution, and receptor specificity";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3060-3064(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92075279; PubMed=1742028;
 RA Berkemeier L.R., Winslow J.W., Kaplan D.R., Nikolics K., Goeddel D.V.,
 RA Rosenthal A.;
 RT "Neurotrophin-5: a novel neurotrophic factor that activates trk and
 trkB";
 RL Neuron 7:857-866(1991).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS).
 RX MEDLINE=20095835; PubMed=10631974;
 RA Robinson R.C., Radziejewski C., Spraggon G., Greenwald J.,
 RA Kostura M.C., Burtnick L.B., Stuart D.I., Choe S., Jones E.Y.;
 RT "The structures of the neurotrophin 4 homodimer and the brain-derived
 neurotrophic factor/neurotrophin 4 heterodimer reveal a common Trk-
 binding site";
 RL Protein Sci. 8:2589-2597(1999).
 CC -!- FUNCTION: TARGET-DERIVED SURVIVAL FACTOR FOR PERIPHERAL SENSORY
 SYMPATHETIC NEURONS.
 CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN PROSTATE, LOWER LEVELS
 IN THYMUS, PLACENTA, AND SKELETAL MUSCLE. EXPRESSED IN EMBRYONIC
 AND ADULT TISSUES.
 CC -!- SIMILARITY: BELONGS TO THE NGF-BETA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC

```

or send an email to license@sib-sib.ch).
-----
CC EMBL; M86528; AAA60154.1; -.
CC DR PIR; JH0503; JH0503.
CC DR PIR; A42687; A42687.
CC DR PDB; 1B8M; 09-FEB-99.
CC DR PDB; 1B98; 26-FEB-99.
CC DR MIM; 162662; -.
CC DR InterPro; IPR002072; NGF.
CC DR Pfam; PF00243; NGF.1.
CC DR PRINTS; PR00268; NGF.
CC DR ProDom; PD002052; NGF.1.
CC DR SMART; SM00140; NGF.1.
CC DR PROSITE; PS00248; NGF.1; 1.
CC DR PROSITE; PS0270; NGF.2; 1.
KW Growth factor; Signal; 3D-structure.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 80
FT CHAIN 81 210 NEUROTROPHIN-5.
FT DISULFID 97 170
FT DISULFID 141 199
FT DISULFID 158 201
FT CARBOHYD 76 76 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 210 AA; 22426 MW; DBC6A30195E139AD CRC64;

Query Match 31.2%; Score 60.5; DB 1; Length 210;
Best Local Similarity 35.0%; Pred. No. 7; 6;
Matches 14; Conservative 3; Mismatches 14; Indels 9; Gaps 1;

OY 3 GPTLRWL-----AARAGGGGGGIEGPTLRQWIA 33
DB 129 GSPLOYFFETRCADNAEEGGPGAGGGCGVDRRHWS 168

RESULT 15
KLT_K_HUMAN
ID KLT_K_HUMAN STANDARD; PRT; 864 AA.
AC P29376;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leukocyte tyrosine kinase receptor precursor (EC 2.7.1.112) (Protein
DE tyrosine kinase-1).
GN LTK OR TYKL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93296146; PubMed=7685902;
RA Toyoshima H., Kozutsumi H., Maru Y., Hagiwara K., Furaya A.,
RA Mtoh H., Hanai N., Takaku F., Yazaki Y., Hirai H.;
RT "Differentially spliced cDNAs of human leukocyte tyrosine kinase
RT receptor tyrosine kinase predict receptor proteins with and without a
RT tyrosine kinase domain and a soluble receptor protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5404-5408(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92007735; PubMed=1655406;
RA Krolewski J.J., Balla-Favera R.;
RT "The ltk gene encodes a novel receptor-type protein tyrosine kinase.";
RL EMBO J. 10:2911-2919(1991).
RN [3]
RP SEQUENCE OF 416-864 FROM N.A.
RX MEDLINE=90206632; PubMed=2320375;
RA Maru Y., Hirai H., Takaku F.;
RT "Human ltk: gene structure and preferential expression in human
RT leukemic cells.";
RL Oncogene Res. 5:199-204(1990).
CC -!- FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS NOT KNOWN. IT IS
CC PROBABLY A RECEPTOR WITH A TYROSINE-PROTEIN KINASE ACTIVITY.

```

```

-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
    tyrosine phosphate.
-!- SUBCELLULAR LOCATION: Type I membrane protein.
-!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; LAMBDA P1, LAMBDA P2
    (SHOWN HERE) AND LAMBDA P3; ARE PRODUCED BY ALTERNATIVE SPLICING.
-!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
    PROTEIN KINASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL; D16105; BAA03679.1; -.
CC EMBL; X60702; CAA43113.1; -.
CC DR EMBL; X52213; CAA36460.1; -.
CC DR PIR; S17452; S17452.
CC DR HSSP; P00523; 2PTK.
CC DR MIM; 151520; -.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002011; Receptor_tyr_kin_II.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; Transmembrane; ATP-binding;
KW Phosphorylation; Receptor; Glycoprotein; Alternative splicing;
KW Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 864 LEUKOCYTE TYROSINE KINASE RECEPTOR.
FT DOMAIN 17 424 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 425 449 POTENTIAL.
FT DOMAIN 450 864 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 510 786 PROTEIN KINASE.
FT NP_BIND 516 524 ATP (BY SIMILARITY).
FT BINDING 544 544 ATP (BY SIMILARITY).
FT ACT_SITE 643 643 BY SIMILARITY.
FT MOD_RES 676 676 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 257 257 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 412 412 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARSPLIC 170 170 G -> VAAASGDGAAPGARAAGPGERAFGLGAGSPAQRG
FT VARSPLIC 171 864 EAPCRFRFPPLPAG (IN ISOFORM LAMBDA P1).
FT VARSPLIC 448 448 MISSING (IN ISOFORM LAMBDA P1).
FT VARSPLIC 449 864 L -> GTKRLAGTVDSLELLSSELGWVSAGSRQ (IN
FT VARSPLIC 449 864 ISOFORM LAMBDA P3).
FT CONFLICT 42 42 MISSING (IN ISOFORM LAMBDA P3).
FT CONFLICT 220 220 Q -> R (IN REF. 2).
FT CONFLICT 274 334 V -> L (IN REF. 2).
FT CONFLICT 449 449 MISSING (IN REF. 2).
FT CONFLICT 452 654 V -> GTKRLAGTVDSLELLSM (IN REF. 3).
FT CONFLICT 652 654 SCA -> MR (IN REF. 3).
SQ SEQUENCE 864 AA; 91653 MW; 97143DD57684A657 CRC64;

Query Match 31.2%; Score 60.5; DB 1; Length 864;
Best Local Similarity 63.6%; Pred. No. 26;
Matches 14; Conservative 1; Mismatches 2; Indels 5; Gaps 2;

OY 2 EG-PTLRWLAAARAGGGGGGG 22
DB 196 EGVPGSRW----AGGGGGGG 213

RESULT 16
JUND_CHICK
ID JUND_CHICK STANDARD; PRT; 323 AA.

```

AC P27921;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Transcription factor Jun-D.
GN JUND.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92019832; PubMed=1923529;
RA Hartl M., Hutchins J.T., Vogt P.K.;
RT "The chicken Jun gene and its product.";
RL Oncogene 6:1623-1631(1991).
CC -1- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. JUN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X60063; CAA42665.1; -
DR PIR: S20099; S20099.
DR HSSP: P05412; IPOS.
DR TRANSFAC: T02196; -
DR InterPro: IPR002112; Leuzip_Jun.
DR InterPro: IPR001871; bZIP.
DR Pfam: PF00170; bZIP.1
DR PRINTS: PR00043; LEUZIPRPUJN.
DR SMART: SM00338; BRLZ.1.
DR PROSITE: PS00036; BZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 59 67 POLY-ALA.
FT DOMAIN 155 166 POLY-GLY.
FT DNA_BIND 242 266 BASIC MOTIF.
FT DOMAIN 270 298 LEUCINE-ZIPPER.
SQ SEQUENCE 323 AA; 33205 MW; A7F6D21A97DBB676 CRC64;

Query Match 30.9%; Score 60; DB 1; Length 323;
Best Local Similarity 72.2%; Pred. No. 12;
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGGIEGPTL 28
Db 151 AAAAGGGGGGGGGGEL 168
||| ||||| |||

RESULT 17
SXL_CERCA STANDARD; PRT; 348 AA.
AC O61374;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sex-lethal protein homolog (CCSL).
DE SXL.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Ceratitis.
OX NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BEAKIO;
RL MEDLINE=98171464; PubMed=9502730;

RA Saccone G., Peluso I., Artiaco D., Giordano E., Bopp D., Polito L.C.;
RT "The Ceratitis capitata homologue of the Drosophila sex-determining
RT gene Sex-lethal is structurally conserved, but not sex-specifically
RL regulated.";
RL Development 125:1495-1500(1998).
CC -1- FUNCTION: UNKNOWN; APPARENTLY NOT INVOLVED IN SOMATIC SEX
CC DETERMINATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS: ADULT-SPECIFIC ISOFORMS
CC A1, A2, A3, A4, AND EMBRYO-SPECIFIC ISOFORMS E1, E2 AND E3 (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYOS OF BOTH SEXES. ALSO
CC EXPRESSED IN THE PROGENITOR CELLS OF THE GERMLINE.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF026145; AAC38968.1; -
DR HSSP: P19339; ISXL.
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; rrm; 2.
DR PRINTS: PR00961; HUDSXLRNA.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS0102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW RNA-binding; Repeat; Nuclear protein; Alternative splicing.
FT DOMAIN 1 27 GLY/ASN-RICH DOMAIN.
FT DOMAIN 110 188 RNA-BINDING (RRM) 1.
FT DOMAIN 196 276 RNA-BINDING (RRM) 2.
FT DOMAIN 68 75 POLY-GLY.
FT DOMAIN 95 99 POLY-GLY.
FT DOMAIN 293 311 POLY-GLY.
FT DOMAIN 312 316 POLY-PRO.
FT VARSPIC 37 44 MISSING (IN ISOFORM A1).
SQ SEQUENCE 348 AA; 37188 MW; CABA3DA5C2C8874A CRC64;

Query Match 30.9%; Score 60; DB 1; Length 348;
Best Local Similarity 83.3%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGP 26
Db 301 GGGGGGGMGGP 312
||||||| |||

RESULT 18
DCO_DROME STANDARD; PRT; 440 AA.
AC O76324;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Discs overgrown protein kinase (EC 2.7.1.-) (Double-time protein).
GN DCO OR DBT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98337188; PubMed=9674431;
RA Kloss B., Price J.L., Saez L., Blau J., Rothenfluh A., Wesley C.S.,
RA Young M.W.;
RT "The Drosophila clock gene double-time encodes a protein closely
RT related to human casein kinase I epsilon.";
RL Cell 94:97-107(1998).

RN MUTAGENESIS, AND FUNCTION.
 RX MEDLINE=98337187; PubMed=9674430;
 RA Price J.L., Blau J., Rothenfluh A., Abodeely M., Kloss B., Young M.W.;
 RT "Double-time is a novel Drosophila clock gene that regulates PERIOD
 protein accumulation.";
 RL Cell 94:83-95(1998).
 CC -!- FUNCTION: INVOLVED IN CIRCADIAN RHYTHMS, VIABILITY AND MOLECULAR
 CC OSCILLATIONS OF THE CLOCK GENES PERIOD (PER) AND TIMELESS (TIM).
 CC DBT REDUCES THE STABILITY AND THUS THE ACCUMULATION OF MONOMERIC
 CC PER PROTEINS, PROBABLY THROUGH PHOSPHORYLATION. NO EVIDENT
 CC CIRCADIAN OSCILLATION IS DETECTED IN HEAD.
 CC -!- SUBUNIT: FORMS A COMPLEX WITH PER.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN PHOTORECEPTOR CELLS OF THE EYES
 CC AS WELL AS IN THE REGION SITUATED BETWEEN THE OPTIC LOBE AND THE
 CC CENTRAL BRAIN.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CASEIN KINASE I SUBFAMILY. COULD BE THE ORTHOLOG OF CKI-EPSILON.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF055583; AAC39134.1; .
 DR HSSP; P40233; 1CSN.
 DR FlyBase; FBgn002413; dco.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 KW Biological rhythms; Transferase; Serine/threonine-protein kinase;
 KW ATP-binding.
 FT DOMAIN 9 277 PROTEIN KINASE.
 FT NP_BIND 15 23 ATP (BY SIMILARITY).
 FT BINDING 38 38 ATP (BY SIMILARITY).
 FT ACT_SITE 128 128 BY SIMILARITY.
 FT DOMAIN 319 332 POLY-ALA.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 351 POLY-GLY.
 FT DOMAIN 414 426 POLY-GLY.
 FT DOMAIN 430 437 POLY-GLY.
 FT MUTAGEN 47 47 P-S; IN DBTS; SHORTENS THE BEHAVIORAL
 FT MUTAGEN 80 80 M->I; IN DBTL; LENGTHENS THE BEHAVIORAL
 FT PERIOD.
 FT SEQUENCE 440 AA; 48073 MW; B875891D5747391D CRC64;
 Query Match 30.9%; Score 60; DB 1; Length 440;
 Best Local Similarity 55.08; Pred. No. 16;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 Qy 4 PTLRWLAARAGGGGGGGI 23
 I : | | | | | | | | | |
 Db 403 PERRPSIRMRGGGGGGGV 422
 RESULT 19
 ID FXD2_HUMAN STANDARD; PRT; 497 AA.
 AC O60548;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Forkhead box protein D2 (Forkhead-related protein FKHL17) (Forkhead-
 DE related transcription factor 9) (FREAC-9).
 GN FOXD2 OR FKHL17 OR FREAC9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98066765; PubMed=9403061;
 RA Ernstsson S., Betz R., Lagercrantz S., Larsson C., Ericksson S.,
 RA Cederberg A., Carlsson P., Enerbaeck S.;
 RT "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
 RT expressed human forkhead gene that maps to chromosome 1p32-p34.";
 RL Genomics 46:78-85(1997).
 CC [2]
 CC REVISIONS.
 CC RA Enerbaeck S.;
 CC Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
 CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF042832; AAC15421.1; .
 DR HSSP; Q63245; 2HFH.
 DR TRANSFAC; T02485; .
 DR MIM; 602211; .
 DR InterPro; IPR001766; Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00039; FORK_HEAD_3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation.
 FT DOMAIN 90 94 POLY-ALA.
 FT DOMAIN 101 104 POLY-ALA.
 FT DNA_BIND 126 217 FORK-HEAD.
 FT DOMAIN 247 250 POLY-ALA.
 FT DOMAIN 296 306 POLY-ALA.
 FT DOMAIN 398 409 POLY-GLY.
 FT DOMAIN 421 426 POLY-GLY.
 FT DOMAIN 442 445 POLY-ALA.
 FT SEQUENCE 497 AA; 49007 MW; EAAAF498D216BE019 CRC64;
 Query Match 30.9%; Score 60; DB 1; Length 497;
 Best Local Similarity 66.7%; Pred. No. 18;
 Matches 14; Conservative 0; Mismatches 5; Indels 2; Gaps 1;
 Qy 4 PT--LRQWLAARAGGGGGGGG 22
 I : | | | | | | | | | |
 Db 385 PTALLRQGLKTDAGGGAGGGG 405
 RESULT 20
 ID K1CJ_MOUSE STANDARD; PRT; 569 AA.
 AC P02535; P08731;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Keratin, type I cytoskeletal 10 (Cytokeratin 10) (56 kDa cytokeratin)
 DE (Keratin, type I cytoskeletal 59 kDa).
 GN KRT10 OR KRT1-10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;


```
DR EMBL; M90470; AAA39451.1; -.
DR EMBL; X52621; CAA36848.1; ALT_SEQ.
DR EMBL; X07984; CAA30793.1; ALT_INIT.
DR PIR; S00904; S00904.
DR PIR; S12792; S12792.
DR HSP; P11362; IFGK.
DR MGD; MGI:96840; Lck.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002011; Receptor_tyr_kin_II.
DR PIR; P00069; pkinase; 1.
DR PRINTS; P00109; TYRKINASE.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; Transmembrane; ATP-binding;
KW Phosphorylation; Receptor; Glycoprotein; Signal; Alternative splicing.
FT SIGNAL 1 16
FT CHAIN 17 888
FT DOMAIN 17 421
FT TRANSMEM 422 446
FT DOMAIN 447 888
FT DOMAIN 506 782
FT NP_BIND 512 520
FT BINDING 540 540
FT ATP (BY SIMILARITY).
FT ACT_SITE 639 639
FT MOD_RES 672 672
FT CARBOHYD 377 377
FT CARBOHYD 409 409
FT VARSPLOC 271 331
FT CONFLICT 789 789
FT CONFLICT 875 875
FT SEQUENCE 888 AA; 94436 MW; 3FFCA80A84863C55 CRC64;

Query Match 30.9%; Score 60; DB 1; Length 888;
Best Local Similarity 63.2%; Pred. No. 30;
Matches 12; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 4 PTLRQLAARAGGGGGGG 22
Db 196 PGWRRW----AGGGGGGG 210

RESULT 22
SUS_DROME
ID SUS_DROME STANDARD; PRT; 1322 AA.
AC P22293;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Suppressor of sable protein.
GN Su(S).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91117256; PubMed=1703632;
RC STRAIN=OREGON-R;
RA Voelker R.A., Gibson W., Graves J.P., Sterling J.F., Eisenberg M.T.;
RT "The Drosophila suppressor of sable gene encodes a polypeptide with
RL regions similar to those of RNA-binding proteins.";
RN Mol. Cell. Biol. 11:894-905(1991).
RN [2]
RP SEQUENCE OF 1-9 FROM N.A.
RX MEDLINE=91169252; PubMed=1963868;
RA Voelker R.A., Graves J.P., Gibson W., Eisenberg M.T.;
RT "Mobile element insertions causing mutations in the prosophila
suppressor of sable locus occur in DNase I hypersensitive subregions
```

```
of 5'-transcribed nontranslated sequences.";
RL Genetics 126:1071-1082(1990).
-!- FUNCTION: AFFECTS THE TRANSCRIPT LEVELS OF THOSE ALLELES THAT IT
CC SUPPRESSES. MAY BE INVOLVED IN RNA METABOLISM.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: AT ALL STAGES.
CC -!- SIMILARITY: HAS REGIONS SIMILAR TO THOSE OF RNA-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57889; AAA28920.1; -.
DR EMBL; X59364; CAA42010.1; -.
DR PIR; A39612; A39612.
DR FlyBase; FBgn0003575; su(s).
DR InterPro; IPR000571; zf-CCCH.
DR Pfam; PF00642; zf-CCCH; 2.
KW RNA-binding; Nuclear protein.
FT DOMAIN 138 327
FT DOMAIN 446 474
FT DOMAIN 1087 1162
FT SEQUENCE 1322 AA; 143555 MW; D5F534EB5702EA08 CRC64;

Query Match 30.9%; Score 60; DB 1; Length 1322;
Best Local Similarity 68.8%; Pred. No. 43;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPTLRQ 30
Db 1159 GGGGGGGVLPNLSQ 1174

RESULT 23
SOX1_MOUSE
ID SOX1_MOUSE STANDARD; PRT; 391 AA.
AC P53783;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SOX-1 protein.
GN SOX1 OR SOX-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=96189340; PubMed=8625802;
RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,
RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,
RA Lovell-Badge R.;
RT "A comparison of the properties of Sox-3 with Sry and two related
RT genes, Sox-1 and Sox-2.";
RL Development 122:509-520(1996).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
CC SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.
CC -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```

DR EMBL; X94126; CAA63846.1; -.
DR HSP; Q05066; 1HRY.
DR MGD; MG1:98357; Sox1.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
KW DNA-binding; Nuclear protein.
FT DOMAIN 30 43 POLY-GLY.
FT DNAS_BIND 51 119 HMG_BOX.
FT DOMAIN 145 150 POLY-GLY.
FT DOMAIN 197 204 POLY-ALA.
FT DOMAIN 280 288 POLY-ALA.
FT DOMAIN 296 306 POLY-ALA.
FT DOMAIN 357 364 POLY-ALA.
SQ SEQUENCE 391 AA; 39237 MW; 9F81ED667F947C05 CRC64;

Query Match 30.7%; Score 59.5; DB 1; Length 391;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 12; Conservative 1; Mismatches 4; Indels 5; Gaps 1;

QY 1 IEPTLRQWLAARAGGGGGG 22
DB 22 LSGPA-----GARGGGGGGG 38

RESULT 24
BET3_MESAU
ID BET3_MESAU STANDARD; PRT; 367 AA.
AC Q09029;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE BET3 protein.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96140430; PubMed=8552091;
RA Peyton M., Stellicrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,
RA Tsai M.-J.;
RT "BETA3, a novel helix-loop-helix protein, can act as a negative
RT regulator of BETA2 and MyoD-responsive genes.";
RL Mol. Cell. Biol. 16:626-633(1996).
CC -!- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3
CC (E47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF
CC NEUROD1 AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION
CC WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE
CC OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.
CC -!- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S08070; AAB50691.1; -.
DR InterPro; IPR003015; HLH_Myc.
DR InterPro; IPR001092; HLH_dlm.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW Nuclear protein; Transcription regulation; Repressor.
FT DOMAIN 11 14 POLY-ALA.
SQ SEQUENCE 11 AA; 14

```

```

FT DOMAIN 58 62 POLY-SER.
FT DOMAIN 83 99 POLY-GLY.
FT DOMAIN 174 179 POLY-GLY.
FT DOMAIN 204 217 POLY-GLY.
FT DNAS_BIND 229 240 BASIC DOMAIN.
FT DOMAIN 241 282 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 311 319 POLY-ALA.
SQ SEQUENCE 367 AA; 35905 MW; 6CAB9AFF9685F77 CRC64;

Query Match 30.4%; Score 59; DB 1; Length 367;
Best Local Similarity 61.1%; Pred. No. 18;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 11 AARAGGGGGGIEGPTL 28
DB 88 AGGGGGGGGGVSVPL 105

RESULT 25
HB9_HUMAN
ID HB9_HUMAN STANDARD; PRT; 401 AA.
AC P50219;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Homeobox protein HB9.
GN HLXB9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94327547; PubMed=7914194;
RA Harrison K.A., Druey K.M., Deguchi Y., Tuscano J.M., Kehrl J.H.;
RA "A novel human homeobox gene distantly related to proboscipedia is
RA expressed in lymphoid and pancreatic tissues.";
RL J. Biol. Chem. 269:19968-19975(1994).
CC -!- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.
CC -!- SIMILARITY: TO DROSOPHILA HOMEOBOX PROTEIN PROBOSCIPEDIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07664; AAB60647.1; -.
DR EMBL; U07663; AAB60647.1; JOINED.
DR HSP; P14653; IB72.
DR TRANSFAC; T03420; -.
DR MIM; 142994; -.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEOBOX.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS00071; HOMEOBOX_2; 1.
KW Homeobox; DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 39 48 POLY-GLY.
FT DOMAIN 97 111 POLY-GLY.
FT DOMAIN 120 135 POLY-ALA.
FT DOMAIN 169 177 POLY-ALA.
FT DNAS_BIND 242 301 HOMEOBOX.
FT DOMAIN 316 325 POLY-GLY.
SQ SEQUENCE 401 AA; 40932 MW; 0006AED71D594FE CRC64;

Query Match 30.4%; Score 59; DB 1; Length 401;

```

```
Best Local Similarity 64.7%; Pred. No. 19;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGGIEGPT 27
   : ||||| | |
Db 37 ASGTGGGGGGGASGCT 53

RESULT 26
ONC2_HUMAN
ID ONC2_HUMAN STANDARD; PRT; 485 AA.
AC O95948;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE One cut domain family member 2 (ONECUT-2 transcription factor) (OC-2).
GN ONECUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99115605; PubMed=9915796;
RA Jacquemin P., Lannoy V., Rousseau G.G., Lemaigre F.P.;
RT "OC-2, a novel mammalian member of the ONECUT class of homeodomain
RT transcription factors whose function in liver partially overlaps with
RT that of hepatocyte nuclear factor-6.";
RL J. Biol. Chem. 274:2665-2671(1999).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES THE TRANSCRIPTION
CC OF A NUMBER OF LIVER GENES SUCH AS HNF3B.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 CUT DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEBOX PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Y18198; CAB38253.1; -
DR TRANSFAC; T03259; -
DR TMW; 604894; -
DR InterPro; IPR003350; CUT.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF02376; CUT; 1.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Activator.
FT DNA_BIND 305 391 CUT.
FT DNA_BIND 407 466 HOMEBOX.
FT DOMAIN 18 37 POLY-GLY.
FT DOMAIN 62 66 POLY-PRO.
FT DOMAIN 75 82 POLY-ALA.
FT DOMAIN 152 165 POLY-HIS.
FT DOMAIN 298 303 POLY-SER.
SQ SEQUENCE 485 AA; 52482 MW; AF21E052EFBE5DA1 CRC64;

Query Match 30.4%; Score 59; DB 1; Length 485;
Best Local Similarity 65.0%; Pred. No. 22;
Matches 13; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPTLRQLAA 34
   : ||||| | |
Db 25 GGGGGGGGGGPGHEQELLA 44

Best Local Similarity 64.7%; Pred. No. 19;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

RESULT 27
ZIN_HUMAN
ID ZIN_HUMAN STANDARD; PRT; 753 AA.
AC O9NRL3;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Zinedin.
GN ZIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347911; PubMed=10748158;
RA Casters F., Rakitina T., Gaillard S., Moqrish A., Mattei M.-G.,
RA Monneron A.;
RT "Zinedin, SG2NA, and striatin are calmodulin-binding, WD repeat
RT proteins principally expressed in the brain.";
RL J. Biol. Chem. 275:19970-19977(2000).
RN [2]
RP SEQUENCE OF 402-753 FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS CALMODULIN IN A CALCIUM DEPENDENT MANNER. MAY
CC FUNCTION AS SCAFFOLDING OR SIGNALING PROTEIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-BOUND (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE STRIATIN FAMILY OF WD-REPEAT PROTEINS.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -!- CAUTION: The name "zinedin" probably originates from the name of
CC the famous soccer player from Marseille (Zinedine Zidane)!
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF212940; AAF29527.1; -
DR EMBL; BC004910; AA04910.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS50082; WD_REPEATS_2; 4.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Calmodulin-binding; Repeat; WD repeat; Coiled coil.
FT DOMAIN 69 136 COILED COIL (POTENTIAL).
FT DOMAIN 165 182 CALMODULIN-BINDING (POTENTIAL).
FT REPEAT 436 475 WD 1.
FT REPEAT 489 528 WD 2.
FT REPEAT 542 581 WD 3.
FT REPEAT 587 628 WD 4.
FT REPEAT 635 674 WD 5.
FT REPEAT 677 716 WD 6.
FT REPEAT 723 752 WD 7.
FT SITE 71 79 CAVEOLIN-BINDING (POTENTIAL).
FT DOMAIN 6 14 POLY-ALA.
FT CONFLICT 402 404 LAD -> GTR (IN REF. 2).
SQ SEQUENCE 753 AA; 80591 MW; 4DA016A8FF7EDB5E CRC64;

Query Match 30.4%; Score 59; DB 1; Length 753;
Best Local Similarity 78.6%; Pred. No. 33;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 AGGGGGGGGIEGPT 27
   : ||||| | |
```

```

Db 44 AGKGGGGGGSPGPT 57
RESULT 28
ECR_LUCCU STANDARD; PRT; 757 AA.
AC 018331;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-ecdysone
DE receptor) (20E receptor).
GN ECR OR NR1H1.
OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Oestroidea; Calliphoridae; Lucilia.
OX NCBI_TaxID=7375;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97449774; PubMed=9304790;
RA Hannan G.N., Hill R.J.;
RT "Cloning and characterization of LcEcR: a functional ecdysone
RT receptor from the sheep blowfly Lucilia cuprina."
RL Insect Biochem. Mol. Biol. 27:479-488(1997).
CC -!- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE
CC ELEMENTS (ECRES) (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U75355; AAB81130.1;
CC HSP; P20393; IAGY.
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR001723; Strdhormone_receptor.
CC InterPro; IPR001628; zf-C4.
CC Pfam; PF00104; hormone_rec. 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR00047; STROIDFINGER.
CC SMART; SM00430; HOLI. 1.
CC SMART; SM00399; ZnF_C4; 1.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW zinc-finger.
FT DOMAIN 1 300 MODULATING (POTENTIAL).
FT DNA_BIND 301 366 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 301 321 C4-TYPE.
FT ZN_FING 337 361 C4-TYPE.
FT DOMAIN 454 674 HORMONE-BINDING (POTENTIAL).
FT SEQUENCE 757 AA; 83075 MW; C1511452ED37D359 CRC64;
SQ
Query Match 30.4%; Score 59; DB 1; Length 757;
Best Local Similarity 76.9%; Pred. No. 33;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 15 GGGGGGGGIEGPT 27
|||||||
Db 129 GGGGGGGGVPGMT 141
RESULT 29
DYHA_CHLRE STANDARD; PRT; 4499 AA.
AC Q39610;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dynein alpha chain, flagellar outer arm (DHC alpha).
GN ODA11 OR ODA-11.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS.
RX STRAIN=21GR;
RX MEDLINE=97329535; PubMed=9186009;
RA Mitchell D.R., Brown K.S.;
RT "Sequence analysis of the Chlamydomonas reinhardtii flagellar alpha
RT dynein gene."
RL Cell Motil. Cytoskeleton 37:120-126(1997).
RN [2]
RP SEQUENCE OF 1142-4499 FROM N.A.
RX STRAIN=21GR;
RX MEDLINE=94274778; PubMed=8006077;
RA Mitchell D.R., Brown K.S.;
RT "Sequence analysis of the Chlamydomonas alpha and beta dynein heavy
RT chain genes."
RL J. Cell Sci. 107:635-644(1994).
CC -!- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND
CC FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES.
CC DYNEIN HAS ATPASE ACTIVITY.
CC -!- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND
CC GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L26049; AAA57316.2;
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001298; Filamin.
CC InterPro; IPR002909; IPT_TIG.
CC InterPro; IPR001798; Kelch.
CC InterPro; IPR001736; PLD.
CC Pfam; PF00630; Filamin; 1.
CC Pfam; PF01344; Kelch; 3.
CC SMART; SM00382; AAA; 3.
CC SMART; SM00429; IPT; 1.
CC PROSITE; PS0194; FILAMIN_REPEAT; 1.
KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
KW Coiled coil.
FT REPEAT 425 534 FILAMIN.
FT DOMAIN 1261 1334 COILED COIL (POTENTIAL).
FT DOMAIN 1382 1450 COILED COIL (POTENTIAL).
FT DOMAIN 1836 1864 MICROTUBULE-BINDING (POTENTIAL).
FT DOMAIN 2655 2688 COILED COIL (POTENTIAL).
FT DOMAIN 3003 3023 COILED COIL (POTENTIAL).
FT DOMAIN 3170 3262 COILED COIL (POTENTIAL).
FT DOMAIN 3486 3515 COILED COIL (POTENTIAL).
FT NP_BIND 1716 1723 ATP (POTENTIAL).
FT NP_BIND 2019 2026 ATP (POTENTIAL).
FT NP_BIND 2369 2376 ATP (POTENTIAL).
FT NP_BIND 2717 2754 ATP (POTENTIAL).
SQ SEQUENCE 4499 AA; 503606 MW; 319AC7FD30F1591A CRC64;
Query Match 30.4%; Score 59; DB 1; Length 4499;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 3 GPTLRWLAAARAGGGGGG 22
|||||

```

Db 4194 GETLFKTVVEVAGGGGGG 4213

RESULT 30

HXD9_HUMAN STANDARD; PRT; 342 AA.

AC P28356;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-JUL-1993 (Rel. 26, Last annotation update)

DE Homeobox protein Hox-D9 (Hox-4C) (Hox-5.2).

GN HOXD9 OR HOX4C.

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Spinal cord;

RX MEDLINE=92097538; PubMed=1756725;

RA Zappavigna V., Renucci A., Izpisua-Belmonte J.-C., Urier G.,

RA Peschle C., Duboule D.;

RT "HOX4 genes encode transcription factors with potential auto- and

RT cross-regulatory capacities.;"

RL EMBL J. 10:4177-4187(1991).

RN [2]

RP SEQUENCE OF 264-342 FROM N.A.

RX MEDLINE=89306602; PubMed=2568311;

RA Oliver G., Sidell N., Fiske N., Heinzmann C., Mohandas T.,

RA Sparkes R.S., de Robertis E.M.;

RT "Complementary homeo protein gradients in developing limb buds.;"

RL Genes Dev. 3:641-650(1989).

RN [3]

RP SEQUENCE OF 275-340 FROM N.A.

RX MEDLINE=90098876; PubMed=2574852;

RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,

RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;

RT "The human HOX gene family.;"

RL Nucleic Acids Res. 17:10385-10402(1989).

CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING LIMB BUDS.

CC -!- SIMILARITY: BELONGS TO THE HOX-B FAMILY OF HOMEBOX PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X59372; CAA42016.1; -

DR EMBL; X15506; CAA33528.1; -

DR PIR; S18649; S18649.

DR PIR; S05958; S05958.

DR PIR; A32830; A32830.

DR HSSP; P02834; 1B81.

DR TRANSFAC; T01424; -

DR MIM; 142982; -

DR InterPro; IPR001356; Homeobox.

DR Pfam; PF00046; homeobox; 1.

DR PRINTS; PR00024; HOMEBOX.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX_1; 1.

DR PROSITE; PS50071; HOMEBOX_2; 1.

KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;

KW Transcription regulation.

FT DOMAIN 115 149 GLY-RICH.

FT DOMAIN 121 130 POLY-GLY.

FT DOMAIN 165 178 SER/THR-RICH.

FT

FT DNA_BIND 275 334 HOMEBOX.

FT CONFLICT 266 E -> A (IN REF. 2).

SQ SEQUENCE 342 AA; 35580 MW; 731981FE25C5ACD7 CRC64;

Query Match 30.2%; Score 58.5; DB 1; Length 342;

Best Local Similarity 44.8%; Pred. No. 19;

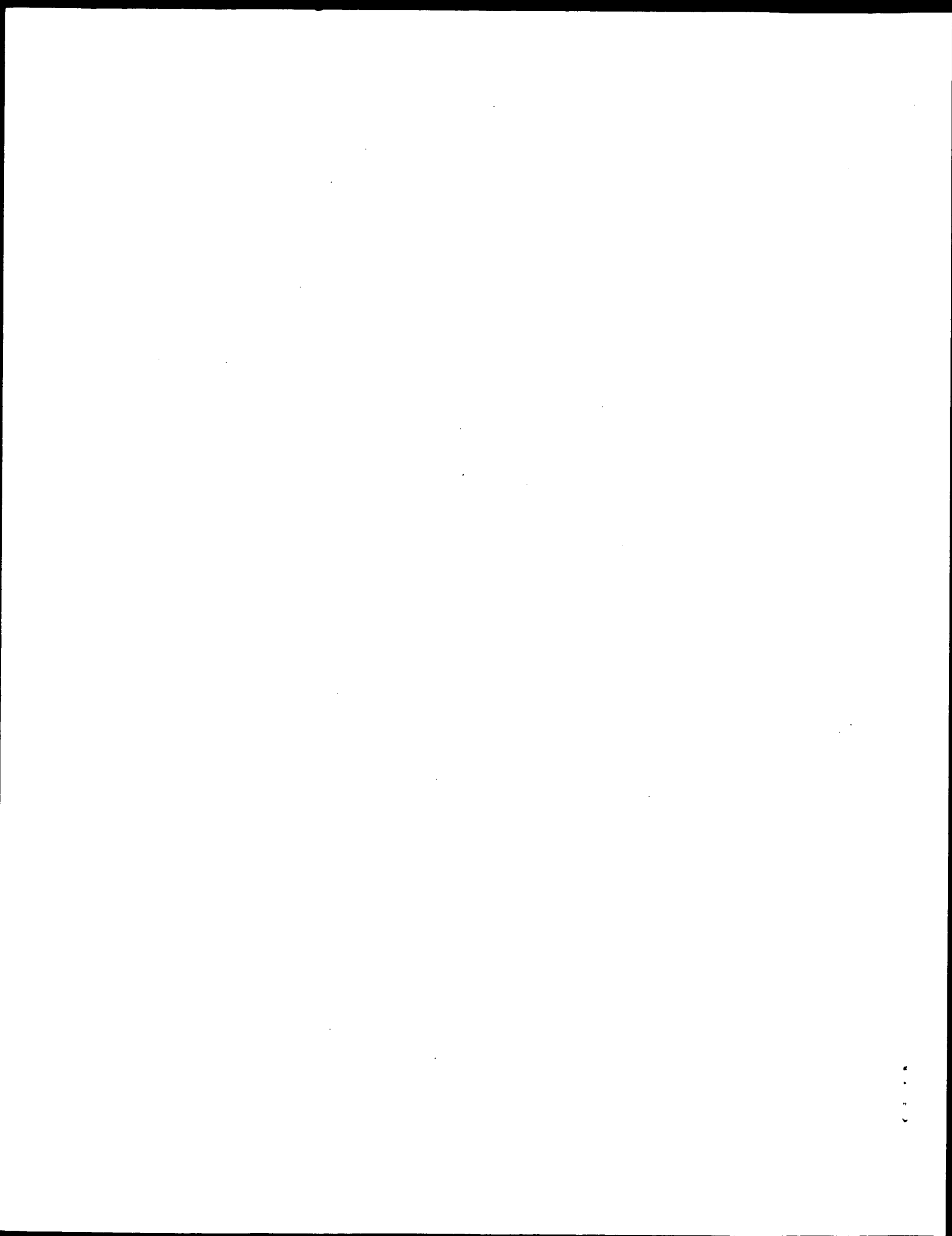
Matches 13; Conservative 2; Mismatches 9; Indels 5; Gaps 1;

QY 3 GPTLRQWLAAARAG-----GGGGGGGIEGP 26

DB 103 GRYVRSWMEPLPGFPGGAGGGGGGGGGP 131

Search completed: October 9, 2002, 09:00:13

Job time : 5.3831 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: October 9, 2002, 08:52:16 ; Search time 12.8993 Seconds
(without alignments)
482.803 Million cell updates/sec

Title: US-09-422-838c-26

Perfect score: 194

Sequence: 1 IEPTLRQLAARAGGGGGGIEPTLRQLAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	74	38.1	360	10 Q9LGC9	Q9LGC9 oryza sativ
2	73.5	37.9	431	13 Q9PVG9	Q9PVG9 coturnix co
3	71	36.6	253	10 Q943K0	Q943K0 oryza sativ
4	70	36.1	439	10 Q9SDK6	Q9SDK6 oryza sativ
5	69	35.6	500	5 Q19476	Q19476 caenorhabdi
6	68.5	35.3	488	16 Q9CCCO	Q9CCCO mycobacteri
7	68.5	35.3	518	2 Q49843	Q49843 mycobacteri
8	68	35.1	125	10 Q9LWC8	Q9LWC8 oryza sativ
9	68	35.1	776	3 Q9HEA4	Q9HEA4 neurospora
10	67	34.5	170	5 Q9W033	Q9W033 drosophila
11	66.5	34.3	202	10 Q9FT25	Q9FT25 oryza sativ
12	66.5	34.3	495	16 Q3230	Q3230 mycobacteri
13	66.5	34.3	496	2 Q9AD76	Q9AD76 streptomyce
14	66	34.0	377	13 Q9YHD0	Q9YHD0 petromyzon
15	66	34.0	529	10 Q9ASE5	Q9ASE5 oryza sativ
16	66	34.0	612	4 Q9P270	Q9P270 homo sapien

17	65.5	33.8	243	10 Q9AR44	Q9ar44 oryza sativ
18	65.5	33.8	1548	4 Q9NYW9	Q9nyw9 homo sapien
19	65.5	33.8	2161	4 Q9Y566	Q9y566 homo sapien
20	65	33.5	447	13 Q73628	Q73628 anolis caro
21	65	33.5	452	5 Q9VJK4	Q9vjk4 drosophila
22	64	33.0	309	5 Q9VV01	Q9vv01 drosophila
23	64	33.0	331	5 Q9U211	Q9u211 caenorhabdi
24	64	33.0	333	5 Q9U210	Q9u210 caenorhabdi
25	64	33.0	422	5 Q96755	Q96755 branchiosto
26	63.5	32.7	207	10 Q94IW9	Q94iw9 oryza sativ
27	63.5	32.7	584	10 Q9LII6	Q9lil6 oryza sativ
28	63	32.5	66	12 Q9LBC5	Q9lbc5 spodoptera
29	63	32.5	137	10 Q9M6A1	Q9m6a1 catharanthu
30	63	32.5	160	10 Q9M699	Q9m699 catharanthu
31	63	32.5	186	10 Q942R8	Q942r8 oryza sativ
32	63	32.5	474	4 Q96SQ2	Q96sq2 homo sapien
33	63	32.5	490	10 Q04270	Q04270 chlamydomon
34	63	32.5	688	4 Q9BYD8	Q9byd8 homo sapien
35	63	32.5	689	4 Q96JG7	Q96jg7 homo sapien
36	63	32.5	752	4 Q96L34	Q96l34 homo sapien
37	63	32.5	841	10 Q9SXI9	Q9sxi9 oryza sativ
38	62.5	32.2	775	4 Q9C0I1	Q9c0i1 homo sapien
39	62	32.0	165	2 Q9AFI5	Q9afi5 mycobacteri
40	62	32.0	286	13 Q9PUX6	Q9pux6 gadus morhu
41	62	32.0	334	11 Q9JKB4	Q9jkb4 mus musculu
42	62	32.0	381	10 Q9LD54	Q9ld54 oryza sativ
43	62	32.0	540	2 Q93H33	Q93h33 streptomyce
44	62	32.0	642	13 Q9PUD8	Q9pud8 lampetra fl
45	62	32.0	664	5 Q9NEC7	Q9nec7 leishmania

ALIGNMENTS

RESULT 1

Q9LGC9 ID Q9LGC9 PRELIMINARY; PRT; 360 AA.
AC Q9LGC9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE PUTATIVE ZINC FINGER PROTEIN.
GN P0462H08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0462H08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002525; BAB07996.1; -
DR InterPro: IPR000571; Zf-CCCH.
DR Pfam: PF00642; Zf-CCCH; 4.
DR SMART: SM00356; Znf_C3H1; 4.
SQ SEQUENCE 360 AA; 37368 MW; 5105598D7E1C77B2 CRC64;

Query Match 38.1%; Score 74; DB 10; Length 360;

Best Local Similarity 56.08; Pred. No. 1.4;

Matches 14; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 IEPTLRQLAARAGGGGGGIEPTLRQLAARA 25

:||| | | | | | | | |

Db 26 LEGPWRMRLGCGGGGGGGGGG 50

RESULT 2

Q9PVG9

ID Q9PVG9 PRELIMINARY; PRT; 431 AA.

Dd 80 GPTVGVYRAGAGGGGPRGFALK 106

RESULT 4
Q9SDK6 PRELIMINARY; PRT; 439 AA.

ID Q9SDK6
AC AC
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartioideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0705D01."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RW EMBL; AP000492; BAA84610.1; -
KW Hypothetical protein.
SQ SEQUENCE 439 AA; 47297 MW; 533EEC240CEA1BA2 CRC64;

Query Match 36.1%; Score 70; DB 10; Length 439;
Best Local Similarity 34.0%; Pred. No. 4.6;
Matches 17; Conservative 2; Mismatches 17; Indels 14; Gaps

QY 1 IEGPLROWLAARAGGGGGG-----TEGPLROWLAARA 36
+ + + + + | | | | | | | | | | : | | | | |
Db 39 LHAPLLPLCGGGGGGGGGGGGGERVGAVGVRGEARSQRAREA 88
+ + + + + | | | | | | | | | | : | | | | |

RESULT 5
Q19476 PRELIMINARY; PRT; 500 AA.

ID Q19476
AC Q19476;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F15B9.5 PROTEIN.
GN F15B9.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SQ SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z78013; CAB01420.1; -
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 500 AA; 53946 MW; 1416327086FE7CF6 CRC64;

Query Match 35.6%; Score 69; DB 5; Length 500;
Best Local Similarity 56.5%; Pred. No. 6.7;
Matches 13; Conservative 4; Mismatches 6; Indels 0; Gaps

QY 3 GPTLROWLAARAGGGGGGTIEG 25
+ + + + + | | | | | | | | | | : | | | | |
Db 429 GSMGRFLSNRGGGGGGGGMGG 451

RESULT 6

Q9CCCO PRELIMINARY; PRT; 488 AA.
 AC Q9CCCO;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
 DE POSSIBLE ATP/GTP-BINDING PROTEIN.
 GN ML0997.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Janczewska C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL583920; CAC31378.1; -.
 DR Leproma; ML0997; -.
 DR InterPro; IPR000765; GTP1_OBG.
 DR PRINTS; PR00326; GTP1_OBG.
 KW Complete proteome.
 SQ SEQUENCE 488 AA; 52800 MW; 188918856F9774AA CRC64;

Query Match 35.3%; Score 68.5; DB 16; Length 488;
 Best Local Similarity 46.7%; Pred. No. 7.4;
 Matches 14; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGGGGGIEGP 26

Db 189 PRLRGWESMSRQVGGGAGGGVGLRGP 218

RESULT 7

Q49843 PRELIMINARY; PRT; 518 AA.
 AC Q49843;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE HFLX.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith D.R.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robison K.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U00019; AAA1774.1; -.
 SQ SEQUENCE 518 AA; 56001 MW; 6641916CC84F374B CRC64;

Query Match 35.3%; Score 68.5; DB 2; Length 518;
 Best Local Similarity 46.7%; Pred. No. 7.8;
 Matches 14; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 4 PTLROW-----LAARAGGGGGGIEGP 26
 Db 219 PRLRGWESMSRQVGGGAGGGVGLRGP 248

RESULT 8

Q9LWC8 PRELIMINARY; PRT; 125 AA.
 AC Q9LWC8;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:PO483F08.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002094; BAA96216.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 125 AA; 13396 MW; C609D8D0B07BC505 CRC64;

Query Match 35.1%; Score 68; DB 10; Length 125;
 Best Local Similarity 42.9%; Pred. No. 2.2;
 Matches 18; Conservative 2; Mismatches 8; Indels 14; Gaps 2;

QY 2 EGPTLROWLAARA-----GGGGGGGGGIEPTLRQ 30

Db 83 EGAAR-WRAARSPARGGQRRHRRGGGGGGGRRPRRR 123

RESULT 9

Q9HEA4 PRELIMINARY; PRT; 776 AA.
 AC Q9HEA4;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN BILAS.200.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL451109; CAC18624.2; -.
 KW Hypothetical protein.
 SQ SEQUENCE 776 AA; 82771 MW; C9BEA870D94A37DE CRC64;

Query Match 35.1%; Score 68; DB 3; Length 776;
 Best Local Similarity 57.7%; Pred. No. 13;
 Matches 15; Conservative 3; Mismatches 4; Indels 4; Gaps 2;

QY 15 GGGGGGGGII--EG-PTLROWLAARA 36

Db 678 GGGGGGGGVVDDDDGPDFAAGLAAQA 703

```

RESULT 10
Q9W033 PRELIMINARY; PRT; 170 AA.
AC Q9W033
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CG13807 PROTEIN.
GN CG13807
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A., Fleischmann W.,
RA Folsler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003474; AAF47627.1; -.
DR FlyBase; FBgn0035323; CG13807.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHL.
SQ SEQUENCE 170 AA; 19099 MW; 477D79D55ADFACES CRC64;

Query Match 34.5%; Score 67; DB 5; Length 170;
Best Local Similarity 50.0%; Pred. No. 3.7;
Matches 12; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 EGPTLRQWLAARAGGGGGGIEG 25
Db 47 EPPIVENWM-----GGGGGGGGGFG 66

RESULT 11
Q9FTZ5 PRELIMINARY; PRT; 202 AA.
AC Q9FTZ5
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE P0436E04.1 PROTEIN.
GN P0436E04.1
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0436E04.1";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002818; BAB16319.1; -.
SQ SEQUENCE 202 AA; 19763 MW; BFC2520037F8E274 CRC64;

Query Match 34.3%; Score 66.5; DB 10; Length 202;
Best Local Similarity 39.0%; Pred. No. 5;
Matches 16; Conservative 5; Mismatches 13; Indels 7; Gaps 1;

QY 1 IEPTLRQWLAARAGGGGG-----GGIEPTLRQWLA 34
Db 94 VVSPSCRQTAGRHGGGGGGRWMAAAGGRDGGGCRWMAA 134

RESULT 12
Q33230 PRELIMINARY; PRT; 495 AA.
AC Q33230
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 53.3 KDA PROTEIN.
GN HFLX OR RV2725C OR MTCY154.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; Z98209; CAB10901.1; -.
DR Tuberculist; Rv2725c; -.
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 495 AA; 53327 MW; F82BA93092945121 CRC64;

Query Match 34.3%; Score 66.5; DB 16; Length 495;
Best Local Similarity 46.7%; Pred. No. 12;
Matches 14; Conservative 1; Mismatches 8; Indels 7; Gaps 1;

QY 4 PTLRQW-----LAARAGGGGGGIEGP 26
Db 199 PRLRWGSMRSQAGGAGGGGGVGLRGP 228

```

```

RESULT 13
Q9AD76 PRELIMINARY; PRT; 496 AA.
AC Q9AD76;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PUTATIVE INTEGRAL MEMBRANE PROTEIN.
GN SCK13.27.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seger K.J., Harris D.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL512667; CAC21636.2; -
DR InterPro; IPR003838; DUF214.
DR Pfam; PF02687; DUF214; 1.
DR QY SEQUENCE 496 AA; 49348 MW; 54E110C4F86231A4 CRC64;

Query Match 34.3%; Score 66.5; DB 2; Length 496;
Best Local Similarity 46.9%; Pred. No. 12;
Matches 15; Conservative 2; Mismatches 6; Indels 9; Gaps 1;

QY 4 PTLQRLAARAGGGG-----GGGGIEGP 26
|||: | |||||
Db 408 PTLQALGGAGGGGGGGGGGGGGGGGGGLGGP 439

RESULT 14
Q9YHDO PRELIMINARY; PRT; 377 AA.
AC Q9YHDO;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 13, Last annotation update)
DE Otx.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RA Tomsa J.M., Langeland J.A.;
RT "Otx expression during lamprey embryogenesis provides insights into
RT the evolution of the vertebrate head and jaw.";
RL Dev. Biol. 0:0-0(1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: WITH OTHER HOMEBOX PROTEINS.
DR EMBL; AF099746; AAC82470.1; -.
DR HSSP; P06601; 1FJL.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.

KW DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 377 AA; 37998 MW; C2DBC19402D3A172 CRC64;

Query Match 34.0%; Score 66; DB 13; Length 377;
Best Local Similarity 48.1%; Pred. No. 11;
Matches 13; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 2 EGPTRLQRLAARAGGGGGGGGGIEGPTL 28
|||: | |||||
Db 265 QGYTAASYGVGEGGGGGGGGGGGPYL 291

RESULT 15
Q9ASE5 PRELIMINARY; PRT; 529 AA.
AC Q9ASE5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE P0456F08.14 PROTEIN.
GN P0456F08.14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:p0456F08.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002901; BAE39414.1; -.
DR InterPro; IPR002937; Amino_oxidase.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF01593; Amino_Oxidase; 1.
DR QY SEQUENCE 529 AA; 55981 MW; 0A5DA55CDD076D24 CRC64;

Query Match 34.0%; Score 66; DB 10; Length 529;
Best Local Similarity 68.4%; Pred. No. 15;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 LQWLAAARAGGGGGGGGIE 24
|||: |||: |||||
Db 151 LRAYQAARSAGGGGGGKE 169

RESULT 16
Q9P270 PRELIMINARY; PRT; 612 AA.
AC Q9P270;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KIAA1458 PROTEIN (FRAGMENT).
GN KIAA1458.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20277482; PubMed-10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
RT genes.XVII.The complete sequences of 100 new cDNA clones from brain
RT which code for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040891; BAA59582.1; -.
DR NON_TER 1
FT 1
SQ SEQUENCE 612 AA; 65593 MW; 9AA4061D21E1E9FD CRC64;

```

```
Query Match 34.0%; Score 66; DB 4; Length 612;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 4 PTLROWLAARAGGGGGGGGIES 25
Db 10 PSLSLRERAGGGGGGGGAG 31

RESULT 17
Q9AR44
ID Q9AR44 PRELIMINARY; PRT; 243 AA.
AC Q9AR44;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE P0498A12.7 PROTEIN (OSJNBA0004B13.18 PROTEIN) (P0581F09.21
DE PROTEIN).
GN P0498A12.7 OR P0581F09.21.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0498A12.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, BAC
RT clone:OSJNBA0004B13.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0581F09.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003020; BAB39979.1;
DR EMBL; AP003018; BAB39964.1;
DR EMBL; AP003631; BAB64224.1;
SQ SEQUENCE 243 AA; 26243 MW; 029E9344C20E0EC8 CRC64;

Query Match 33.8%; Score 65.5; DB 10; Length 243;
Best Local Similarity 52.2%; Pred. No. 7;
Matches 12; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

Qy 1 IEGPTLROWLAARAGGGGGGGI 23
Db 28 VRTPDQRRW---RRGGGGGGGV 47

RESULT 18
Q9NYW9
ID Q9NYW9 PRELIMINARY; PRT; 1548 AA.
AC Q9NYW9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SOMATOSTATIN RECEPTOR-INTERACTING PROTEIN SPLICE VARIANT B.
GN SSTRIIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zitzer H., Hoenck H.-H., Richter D., Kreienkamp H.-J.;
RT "The somatostatin receptor interacting protein (SSTRIP) defines a
RT novel family of multidomain postsynaptic density proteins.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF163302; AAD45121.1;
DR HSSP; P06241; 1SHF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; Receptor; Repeat.
SQ SEQUENCE 2161 AA; 225019 MW; 5FEFC969CBE98701 CRC64;

Query Match 33.8%; Score 65.5; DB 4; Length 2161;
Best Local Similarity 38.1%; Pred. No. 68;
Matches 16; Conservative 2; Mismatches 5; Indels 19; Gaps 2;

Qy 4 PTLROWLAARAGG-----GGGGGGIEGPTLR 29
Db 432 PSLRGW---RGGGPTPGAPSPSHHSGAGGGGGSGGPALR 470

RESULT 19
Q9Y566
ID Q9Y566 PRELIMINARY; PRT; 2161 AA.
AC Q9Y566;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SOMATOSTATIN RECEPTOR INTERACTING PROTEIN SPLICE VARIANT A.
GN SSTRIIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zitzer H., Hoenck H.-H., Richter D., Kreienkamp H.-J.;
RT "The somatostatin receptor interacting protein (SSTRIP) defines a
RT novel family of multidomain postsynaptic density proteins.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF163302; AAD45121.1;
DR HSSP; P06241; 1SHF.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; Receptor; Repeat.
SQ SEQUENCE 2161 AA; 225019 MW; 5FEFC969CBE98701 CRC64;

Query Match 33.8%; Score 65.5; DB 4; Length 1548;
Best Local Similarity 38.1%; Pred. No. 49;
Matches 16; Conservative 2; Mismatches 5; Indels 19; Gaps 2;

Qy 4 PTLROWLAARAGG-----GGGGGGIEGPTLR 29
Db 432 PSLRGW---RGGGPTPGAPSPSHHSGAGGGGGSGGPALR 470
```

Db 1045 PSLRGW---RGGGSPPTGAPSPSHGSGAGGGGSSGCPALR 1083
||||| :||| ||

RESULT 20
O73628 PRELIMINARY; PRT; 447 AA.
AC O73628
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BRAIN-2 GENE.
GN BRAIN-2 GENE.
OS Anolis carolinensis (Green anole) (American chameleon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Iguania; Iguanidae; Polychrotidae; Anolis.
OX NCBI_TaxID=28377;
FN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97475689; PubMed=9335144;
RA Nakachi Y., Hayakawa T., Oota H., Sumiyama K., Wang L., Ueda S.;
RT "Nucleotide compositional constraints on genomes generate alanine-,
RT glycine-, and proline-rich structures in transcription factors.";
RL Mol. Biol. Evol. 14:1042-1049(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: WITH OTHER HOMEBOX PROTEINS.
DR EMBL; AB001869; BAA28666.1; -;
DR HSSP; P14859; 10CT.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000327; POU.
DR Pfam; PF00046; homeobox; 1.
DR Pfam; PF00157; pou; 1.
DR PRINTS; PF00028; POU00028.
DR PRODOM; PD000583; POU; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00352; POU; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PROSITE; PS00035; POU_1; 1.
DR PROSITE; PS00465; POU_2; 1.
KW Brain; DNA-binding; Homeobox; Nuclear protein.
SQ SEQUENCE 447 AA; 47160 MW; AFA362894FCBC419 CRC64;

Query Match 33.5%; Score 65; DB 13; Length 447;
Best Local Similarity 73.3%; Pred. No. 16;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 8 QWLAARAGGGGGGG 22
||| : |||||
Db 56 QWIAALSHGGGGGG 70

RESULT 21
Q9VJK4 PRELIMINARY; PRT; 452 AA.
AC Q9VJK4
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CGS953 PROTEIN.
GN CGS953.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
FN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Fostel C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harkis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003651; AAF53541.1; -;
DR FlyBase; FBgn0032587; CG5953.
SQ SEQUENCE 452 AA; 47875 MW; 0F7ABD53014E3E5C CRC64;

Query Match 33.5%; Score 65; DB 5; Length 452;
Best Local Similarity 73.3%; Pred. No. 16;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 12 ARAGGGGGGGGIEGP 26
| ||||| : ||
Db 259 AAGGGGGGGGVVGP 273

RESULT 22
Q9VV01 PRELIMINARY; PRT; 309 AA.
AC Q9VV01
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CG13055 PROTEIN.
GN CG13055.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
FN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benon P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster."
 RT Science 287:2185-2195(2000).
 RL EMBL: AE003528; AAF49521.1; -
 DR FlyBase: FBgn0036583; CG13055.
 SQ SEQUENCE 309 AA; 33224 MW; 9DAEB67784852A93 CRC64;

Query Match 33.0%; Score 64; DB 5; Length 309;

Best Local Similarity 57.9%; Pred. No. 14;

Matches 11; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 12 ARAGGGGGGGGIEGPTLRQ 30

Db 94 SRSSGGGGGGVAGVTLQ 112

RESULT 23

Q9U211

ID Q9U211 PRELIMINARY; PRT; 331 AA.

AC Q9U211;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE Y41C4A.4A PROTEIN.

GN Y41C4A.4A.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RA SEQUENCE FROM N.A.

RA Steward C.A.

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

investigating biology."

RL Science 282:2012-2018(1998).

CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.

DR EMBL; AL032627; CAB54381.1; -

DR InterPro; IPR001871; bZIP.

DR InterPro; IPR003102; pKID.

DR Pfam; PF00170; bZIP; 1.
 DR Pfam; PF02173; pKID; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS00036; BZIP_BASIC; 1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 331 AA; 34985 MW; A414C19D4ADCC91E CRC64;

Query Match 33.0%; Score 64; DB 5; Length 331;
 Best Local Similarity 76.9%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27

Db 167 GGGGGGGGVPGPS 179

RESULT 24

Q9U210

ID Q9U210 PRELIMINARY; PRT; 333 AA.

AC Q9U210;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE Y41C4A.4B PROTEIN.

GN Y41C4A.4B.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Steward C.A.

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for

investigating biology."

RL Science 282:2012-2018(1998).

CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.

DR EMBL; AL032627; CAB54382.1; -

DR InterPro; IPR001871; bZIP.

DR InterPro; IPR003102; pKID.

DR Pfam; PF00170; bZIP; 1.

DR Pfam; PF02173; pKID; 1.

DR SMART; SM00338; BRLZ; 1.

DR PROSITE; PS00036; BZIP_BASIC; 1.

DR DNA-binding; Nuclear protein.

SQ SEQUENCE 333 AA; 35261 MW; BF02CE56398F6D058 CRC64;

Query Match 33.0%; Score 64; DB 5; Length 333;

Best Local Similarity 76.9%; Pred. No. 15;

Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27

Db 169 GGGGGGGGVPGPS 181

RESULT 25

O96755

ID O96755 PRELIMINARY; PRT; 422 AA.

AC O96755;

DT 01-MAY-1999 (Tremblrel. 10, Created)

DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE INTERMEDIATE FILAMENT PROTEIN E1.

OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).

OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

OC Branchiostoma.

OX NCBI_TaxID=7740;

```

RN  SEQUENCE FROM N.A.
RP  MEDLINE-99019308; PubMed-9804163;
RA  Karabinos A., Riener D., Erber A., Weber K.;
RT  "Homologues of vertebrate type I, II and III intermediate filament
RT  (IF) proteins in an invertebrate: the IF multigene family of the
RT  cephalochordate Branchiostoma.";
RL  FEBS Lett. 437:15-18(1998).
DR  EMBL; AJ010294; CAA09068.1; -.
DR  InterPro; IPR002952; Eggshell.
DR  InterPro; IPR001664; IF.
DR  InterPro; IPR002957; Keratin.I.
DR  Pfam; PF0003489; Ribosomal_S30.
DR  PRINTS; PR01228; EGGSHELL.
DR  PRINTS; PR01248; TYPEIKERATIN.
SQ  SEQUENCE 422 AA; 44892 MW; 85FE742F07751B24 CRC64;

Query Match      33.0%; Score 64; DB 5; Length 422;
Best Local Similarity 61.9%; Pred. No. 19;
Matches 13; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

QY  15 GGGGGGGGIEG-----PTLR 29
    |||||
DB  92 GGGGGGGGIGMWTETPTMR 112

RESULT 26
Q94IW9  PRELIMINARY; PRT; 207 AA.
ID  Q94IW9
AC  Q94IW9
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE  P0037C04.13 PROTEIN.
DE  P0037C04.13 PROTEIN.
GN  P0037C04.13.
OS  Oryza sativa (Rice).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC  Ehrhartoideae; Oryzeae; Oryza.
OX  NCBI_TaxID=4530;
RN  SEQUENCE FROM N.A.
RP  STRAIN=CV. NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT  clone: P0037C04.";
RL  Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF003233; BAB5526.1; -.
SQ  SEQUENCE 207 AA; 21266 MW; F514ABC36A6DC403 CRC64;

Query Match      32.7%; Score 63.5; DB 10; Length 207;
Best Local Similarity 45.5%; Pred. No. 11;
Matches 15; Conservative 4; Mismatches 5; Indels 9; Gaps 2;

QY  11 AARAGGGG-----GGGIEGPTLR0WLAARA 36
    |:|||||
DB  122 AVQAGGGGGSQAVQAGG--GGAVRQWCASES 152

RESULT 27
Q9LI16  PRELIMINARY; PRT; 584 AA.
ID  Q9LI16
AC  Q9LI16
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  HYPOTHETICAL PROTEIN.
OS  Oryza sativa (Rice).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC  Ehrhartoideae; Oryzeae; Oryza.
OX  NCBI_TaxID=4530;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=CV. NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT  clone: P0708G02.";
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF001539; BAA92923.1; -.
DR  HSP; P00950; 5PGM.
DR  InterPro; IPR001345; PG_mutase.
DR  Pfam; PF00300; PGAM; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 584 AA; 63515 MW; 351C684C8BBD9CF CRC64;

Query Match      32.7%; Score 63.5; DB 10; Length 584;
Best Local Similarity 48.3%; Pred. No. 30;
Matches 14; Conservative 2; Mismatches 8; Indels 5; Gaps 1;

QY  7 ROWLAARA-----GGGGGGGIEGPTLRQ 30
    | | | | |
DB  113 RWTATRSSDPGIGGGGGGGGAPTRRR 141

RESULT 28
Q91BC5  PRELIMINARY; PRT; 66 AA.
ID  Q91BC5
AC  Q91BC5
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  HYPOTHETICAL 7.0 KDA PROTEIN.
OS  Spodoptera litura nucleopolyhedrovirus.
OC  Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC  Nucleopolyhedrovirus.
OX  NCBI_TaxID=46242;
RN  SEQUENCE FROM N.A.
RP  STRAIN=G2;
RA  Pang Y., Yu J., Wang L., Hu X., Bao W., Li G., Chen C., Han H., Hu S.,
RA  Yang H.;
RT  "Sequence Analysis of the Spodoptera litura Multicapsid
RT  Nucleopolyhedrovirus Genome.";
RL  Virology 287:391-404(2001).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=G2;
RA  Yu J., Wang L., Hu X., Pang Y.;
RL  Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF325155; AAL01786.1; -.
KW  Hypothetical protein.
SQ  SEQUENCE 66 AA; 6998 MW; C5626A8FFA9C9E7C CRC64;

Query Match      32.5%; Score 63; DB 12; Length 66;
Best Local Similarity 68.8%; Pred. No. 3.9;
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY  13 RAGGGGGGGGIEGPTL 28
    |:|||||
DB  19 RSGGGGGGGVVGAML 34

RESULT 29
Q9M6A1  PRELIMINARY; PRT; 137 AA.
ID  Q9M6A1
AC  Q9M6A1
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  PUTATIVE GLYCINE-RICH RNA BINDING PROTEIN 1.
GN  GRP-1.
OS  Catharanthus roseus (rosy periwinkle) (Madagascar periwinkle).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Wed Oct 9 10:29:49 2002

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 OC Vincaeae; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Clastre M., Chenieux J.C., Rideau M., Hamdi S.;
 RT "Genes encoding glycine-rich Catharanthus roseus proteins with RNA-
 binding motifs".
 RL Submitted (Oct-1999) to the EMBL/GenBank/DBDJ databases.

DR EMBL; AF200321; AAF31402.1; -;
 DR HSSP; P09651; 1HA1.
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 SQ SEQUENCE 137 AA; 14162 MW; 4FABADB9C7A989FC CRC64;

Query Match 32.5%; Score 63; DB 10; Length 137;
 Best Local Similarity 50.0%; Pred. No. 8.1;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 TLQWLAAAGGGGGGGIEGP 26
 : : : ||||| ||
 Db 80 TVNEAQRSGGGGGGGFRGP 101

RESULT 30

Q9M699 PRELIMINARY; PRT; 160 AA.
 ID Q9M699
 AC Q9M699;
 DT 01-OCT-2000 (TREMREL. 15, Created)
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE PUTATIVE GLYCINE-RICH RNA-BINDING PROTEIN 2.
 GN GRP-2.
 OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 OC Vincaeae; Catharanthus.
 OX NCBI_TaxID=4058;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veau B., Oudin A., Courtois M., Chenieux J.-C., Hamdi S., Rideau M.,
 RA Clastre M.;
 RT "Cloning of two cDNAs encoding crGRP2 and crGRP3 (Accession Nos.
 RT AF200323 and AF200322), the first members of the RRM-GRP family in
 RT Catharanthus roseus (PGR00-049).";
 RL Plant Physiol. 122:1459-1459(2000).
 DR EMBL; AF200323; AAF31404.1; -;
 DR HSSP; P09651; 1HA1.
 DR InterPro; IPR002952; Eggshell.
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; rrm; 1.
 DR PRINTS; PR01228; EGGSELL.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 SQ SEQUENCE 160 AA; 16264 MW; DDC9F63C983F5F2 CRC64;

Query Match 32.5%; Score 63; DB 10; Length 160;
 Best Local Similarity 50.0%; Pred. No. 9.4;
 Matches 11; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 TLQWLAAAGGGGGGGIEGP 26
 : : : ||||| ||
 Db 80 TVNEAQRSGGGGGGGFRGP 101

Search completed: October 9, 2002, 09:03:08
 Job time: 13.9826 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:50:51 : Search time 16.1874 Seconds
(without alignments)
247.023 Million cell updates/sec

Title: US-09-422-838c-27

Perfect score: 190

Sequence: 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	190	100.0	36	21	AA17298
2	190	100.0	36	21	AA17299
3	190	100.0	36	21	AA17300
4	172	90.5	36	21	AA17301
5	172	90.5	36	21	AA17302
6	168	88.4	36	21	AA17303
7	168	88.4	36	21	AA17304
8	168	88.4	36	21	AA17305
9	168	88.4	41	21	AA17306
10	168	88.4	42	21	AA17307
11	168	88.4	42	21	AA17308

12	168	88.4	42	21	AA17309	Synthetic TMP-TMP
13	168	88.4	42	21	AA17310	Thrombopoietin mim
14	168	88.4	60	21	AA17311	Synthetic TMP-TMP
15	168	88.4	269	21	AA17312	Thrombopoietin mim
16	168	88.4	269	21	AA17313	Thrombopoietin mim
17	164	86.3	288	21	AA17314	Human IgG1 Fc TNP
18	160	84.2	36	21	AA17315	FC-TMP-TMP protein
19	160	84.2	36	21	AA17316	TPO-mimetic peptid
20	159	83.7	36	21	AA17317	Thrombopoietin mim
21	159	83.7	36	21	AA17318	TPO-mimetic peptid
22	159	83.7	36	21	AA17319	Thrombopoietin mim
23	157.5	82.9	37	21	AA17320	TPO-mimetic peptid
24	157	82.6	38	21	AA17321	TPO-mimetic peptid
25	156.5	82.4	39	21	AA17322	TPO-mimetic peptid
26	156.5	82.4	39	21	AA17323	TPO-mimetic peptid
27	156	82.1	36	21	AA17324	TPO-mimetic peptid
28	156	82.1	36	21	AA17325	Thrombopoietin mim
29	155	81.6	42	21	AA17326	TPO-mimetic peptid
30	151.5	79.7	35	21	AA17327	TPO-mimetic peptid
31	148	77.9	40	21	AA17328	TPO-mimetic peptid
32	145	76.3	34	21	AA17329	TPO-mimetic peptid
33	138.5	72.9	33	21	AA17330	TPO-mimetic peptid
34	132	69.5	32	21	AA17331	TPO-mimetic peptid
35	127.5	67.1	29	21	AA17332	TPO-mimetic peptid
36	125.5	66.1	31	21	AA17333	TPO-mimetic peptid
37	119	62.6	30	21	AA17334	TPO-mimetic peptid
38	118	62.1	32	21	AA17335	TPO-mimetic peptid
39	118	62.1	32	21	AA17336	Thrombopoietin mim
40	118	62.1	34	21	AA17337	Thrombopoietin mim
41	116.5	61.3	29	21	AA17338	TPO-mimetic peptid
42	116.5	61.3	29	21	AA17339	TPO-mimetic peptid
43	112.5	59.2	29	21	AA17340	TPO-mimetic peptid
44	106	55.8	28	21	AA17341	TPO-mimetic peptid
45	105.5	55.5	29	21	AA17342	TPO-mimetic peptid

ALIGNMENTS

RESULT 1
AA17298
ID AA17298 standard; Peptide; 36 AA.
XX AA17298;
XX AC
XX 31-OCT-2000 (first entry)
XX DT
XX DE
XX TPO-mimetic peptide sequence SEQ ID NO:354.
XX DE
XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX OS
XX Synthetic.
XX WO200024782-A2.
XX PN
XX 04-MAY-2000.
XX PD
XX TPO-mimetic peptid
XX PF
XX Cyclic or linear t
XX PR
XX TPO-mimetic peptid
XX PR
XX Linear thrombopoie
XX PR
XX TPO-mimetic peptid
XX PR
XX TPO-mimetic peptid
XX PA
XX Thrombopoietin mim
XX PI
XX Thrombopoietin mim
XX BO
XX WPI; 2000-350702/30.

Feige U, Liu C, Cheetham J, Boone TC;

pharmacologically active peptides, useful for treating cancer and autoimmune diseases -

Example 1; Page 318; 608pp; English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

XX Sequence 36 AA;

Query Match 88.4%; Score 168; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 1.6e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36

RESULT 8

AA96525
ID AAY96525 standard; peptide: 36 AA.

XX AAY96525;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 6.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.

XX Synthetic.

Key	Location/Qualifiers
Modified-site	1
Peptide	/note= "optionally linked to an Fc molecule"
Peptide	1..14
Peptide	/label= TMP_1
Peptide	15..18
Peptide	/label= linker
Peptide	19..32
Peptide	/label= TMP_2
Modified-site	32
	/note= "optionally linked to an Fc molecule"

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

PI Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thrombocytopenia
XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L1)-TMP_2],
CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least
CC 10 to 14 residues in length comprising X2-X1-0, X2-X1-1, X2-X1-2,
CC X2-X1-3, X2-X1-4, X1-X1-0, X1-X1-1, X1-X1-2, X1-X1-3, and
CC X1-X1-4. X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
CC X4 = P; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
CC or E; X9 = W, Y or F; X10 = L, I, V, A, F, M, or K; X11 = A, I, V,
CC L, F, S, T, K, H, or E; X12 = A, I, V, L, F, G, S, or Q; X13 = R, K,
CC T, V, N, Q or G; X14 = A, I, V, L, F, T, R, E, or G; L1 = linker
CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
CC activate the c-Mpl receptor which mediates the activity of endogenous
CC thrombopoietin. The TMPs are useful for increasing the production of
CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 36 AA;

Query Match 88.4%; Score 168; DB 21; Length 36;
Best Local Similarity 94.4%; Pred. No. 1.6e-13;
Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36
||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 IEGPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36

RESULT 9

AA96528
ID AAY96528 standard; peptide: 41 AA.

XX AAY96528;

XX 04-SEP-2000 (first entry)

XX Thrombopoietin mimetic peptide compound 9.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker.

XX Synthetic.

Key	Location/Qualifiers
Modified-site	1
Peptide	/note= "optionally linked to an Fc molecule"
Peptide	6..19
Peptide	/label= TMP_1
Peptide	20..27
Peptide	/label= linker
Peptide	28..41
Peptide	/label= TMP_2

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAG59443
 CC to AAG69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 269 AA;

Query Match 88.4%; Score 168; DB 21; Length 269;
 Best Local Similarity 94.4%; Pred. No. 1.2e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36
 DB 2 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 37

RESULT 16
 AAB169531
 ID AAY96531 standard; Protein; 269 AA.

XX AAY96531;
 XX
 DT 04-SEP-2000 (first entry)
 DE Human IgG1 Fc TNP fusion protein.

XX Immunoglobulin; IgG1; Fc; thrombopoietin; mimetic; TNP; TPO; platelet;
 KW megakaryocyte; production; anti-human immunodeficiency virus; anti-HIV;
 KW anti-anemic; dermatological; immunosuppressive; anti-inflammatory.

OS Homo sapiens.

XX WO200024770-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24834.

XX 23-OCT-1998; 98US-0105348.

XX (AMGE-) AMGEN INC.

XX Liu C, Feige U, Cheetham J;

XX WPI; 2000-365108/31.

DR N-PSDB; AAA29229.

XX Thrombopoietic peptides which activate mpl receptors and increase the
 PT production of platelets or platelet precursors, useful for treatment of
 PT diseases which involve thrombocytopenia

XX Example 2a; Page 49-50; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
 CC mimetic peptide (X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAG59443
 CC to AAG69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

CC X1-X1.4. X1 = I, A, V, L, S or R; X2 = E, D, K or V; X3 = G or A;
 CC X4 = P; X5 = T or S; X6 = L, I, V, A or F; X7 = R or K; X8 = Q, N,
 CC or E; X9 = W, Y or F; X10 = L, I, V, A, F, M, or K; X11 = A, I, V,
 CC L, F, S, T, K, H, or E; X12 = A, I, V, L, F, G, S, or Q; X13 = R, K,
 CC T, V, N, Q or G; X14 = A, I, V, L, F, T, R, E, or G; L1 = linker
 CC comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and
 CC activate the c-Mpl receptor which mediates the activity of endogenous
 CC thrombopoietin. The TNP are useful for increasing the production of
 CC platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which
 CC is useful for treatment of diseases which involve thrombocytopenia, e.g.
 CC aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency
 CC virus associated ITP, and systemic lupus erythematosus.

XX Sequence 269 AA;

Query Match 88.4%; Score 168; DB 21; Length 269;
 Best Local Similarity 94.4%; Pred. No. 1.2e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 36
 DB 234 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAARA 269

RESULT 17

AAB16959
 ID AAB16959 standard; Protein; 268 AA.

XX AAB16959;

XX 31-OCT-2000 (first entry)

XX Fc-TNP-TMP protein sequence SEQ ID NO:8.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Homo sapiens.

OS Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

DR N-PSDB; AAA69445.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 2; Page 182-183; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-P1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 268 AA;
Query Match 86.3%; Score 164; DB 21; Length 268;
Best Local Similarity 94.3%; Pred. No. 3.6e-12;
Matches 33; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 35
||||||| ||||||| ||||||| ||||||| |||||
DB 234 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 268

RESULT 18
AAB17301
ID AAB17301 standard; Peptide; 36 AA.

XX
AC AAB17301;

XX
DT 31-OCT-2000 (first entry)

XX
DE TPO-mimetic peptide sequence SEQ ID NO:357.

XX Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTA44; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.

XX
OS Synthetic.

XX
PN WO200024782-A2.

XX
PD 04-MAY-2000.

XX
PF 25-OCT-1999; 99WO-US25044.

XX
PR 23-OCT-1998; 98US-0105371.

XX
PR 22-OCT-1999; 99US-0428082.

XX
PA (AMGE-) AMGEN INC.

XX
PI Feige U, Liu C, Cheetham J, Boone TC;

XX
DR WPI; 2000-350702/30.

PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -

XX
PS Example 1; Page 321; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 = are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein
CC A binding, complement fixation, and possibly placental transfer. AAA69443
CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
CC sequences used in the exemplification of the present invention.

XX
SQ Sequence 36 AA;

Query Match 84.2%; Score 160; DB 21; Length 36;
Best Local Similarity 91.7%; Pred. No. 1.5e-12;
Matches 33; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 36

||||||| ||||||| ||||||| ||||||| |||||

DB 1 IEPTLRQCLAAAGGGGGGIEGPTLRQCLAAAR 36

RESULT 19

AAY96523
ID AAY96523 standard; peptide; 36 AA.

XX
AC AAY96523;

XX
DT 04-SEP-2000 (first entry)

XX
DE Thrombopoietin mimetic peptide compound 4.

XX Thrombopoietin; mimetic; TMP; TPO; platelet; megakaryocyte; production;
KW anti-human immunodeficiency virus; anti-HIV; anti-anemic; dermatological;
KW immunosuppressive; anti-inflammatory; linker; cyclic; linear.

XX
OS Synthetic.

XX
FH Key Location/Qualifiers

FT Modified-site 1 /note= "optionally linked to an Fc molecule"

FT Peptide 1..14

FT Peptide 15..22 /label= TMP_1

FT Modified-site 18 /label= linker

FT Peptide 23..36 /note= "optionally modified by bromoacetyl or PEG"

FT Peptide 23..36 /label= TMP_2

XX
PN WO200024770-A2.

XX
PD 04-MAY-2000.

XX
PF 22-OCT-1999; 99WO-US24834.

XX
PR 23-OCT-1998; 98US-0105348.

XX
PA (AMGE-) AMGEN INC.

XX
PI Liu C, Feige U, Cheetham J;

XX
DR WPI; 2000-365108/31.

XX Thrombopoietic peptides which activate mpl receptors and increase the
PT production of platelets or platelet precursors, useful for treatment of
PT diseases which involve thrombocytopenia
XX Claim 16; Page 62; 91pp; English.

XX A compound which binds to an mpl receptor comprising a thrombopoietin
CC mimetic peptide (TMP) dimer joined by a linker [TMP_1-(L_1)-TMP_2],
CC is new. TMP_1 and TMP_2 are amino acid sequences varying from at least

pharmacologically active peptides; L1, L2, L3, and L4 = are each independently linkers; and a, b, c, d, e, and f = are each independently 0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AA69443 to AA69526 and AAB36955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

```

Query Match      83.7%  Score 159;  DB 21;  Length 36;
Best Local Similarity 91.7%  Pred. NO. 2e-12;
Matches 33;  Conservative 0;  Mismatches 3;  Indels 0;  Gaps 0;
QY  1  IEGPTLRCLAARAGGGGGGIEGPTLRCLAA 36
      |||||  |||||  |||||  |||||  |||||
Db  1  IEGPTLRCLAARAGGGGGGIEGPTLRCLAA 36
      |||||  |||||  |||||  |||||  |||||

```

AAB17307
 ID AAB17307 standard; Peptide; 36 AA.
 XX
 AC AAB17307;
 XX
 XX 31-OCT-2000 (first entry)
 DT
 XX
 DE TPO-mimetic peptide sequence SEQ ID NO:363.

XX	PN	WO200024782-A2.
XX	XX	
XX	PD	04-MAY-2000.
XX	XX	
XX	PF	25-OCT-1999; 99WO-US25044.
XX	XX	
XX	PR	23-OCT-1998; 98US-0105371.
XX	PR	22-OCT-1999; 99US-0428082.
XX	XX	
XX	PA	(AMGE-) AMGEN INC.
XX	XX	
XX	PI	Feige U, Liu C, Cheetham J, Boone TC;
XX	XX	
XX	DR	WPI; 2000-350702/30.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)-a-F1-(X2)_b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from - (L1)-c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)-e-P³, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)-f-P4 where P1, P2, P3, and P4 = are each independently sequences of pharmacologically active peptides; L1, L2, L3, and L4 = are each pharmacologically linkers; and a, b, c, d, e, and f = are each independently

0 or 1, provided that at least 1 of a and b is 1. The composition can have cytostatic, antiasthmatic, thrombolytic and immunosuppressive activities. DNAs, vectors and host cells from the present invention can be used for producing pharmaceutical compositions. The compositions are useful for treating cancer, asthma, thrombosis, or autoimmune diseases. The use of an Fc domain (rather than a Fab domain) can provide a longer half-life or incorporate functions such as Fc receptor binding, protein A binding, complement fixation, and possibly placental transfer. AAA69443 to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid sequences used in the exemplification of the present invention.

Seq	Sequence	36 AA;
Query Match	83.7%	Score 159; DB 21; Length 36;
Best Local Similarity	91.7%	Pred. NO. 2e-12;
Matches	33; Conservative	0: Mismatches 3: Indels 0: Gaps 0:

Qy 1 IEGPTLRQCLAAARCGGGGGGIEGPTLRQCLAARA 36
| | | | | | | | | | | | | | | | | |
Dh 1 IEGPTLRQCLAAARCGGGGGGIEGPTLRQCLAARA 36
| | | | | | | | | | | | | | | | | |

RESULT 22
AAAY96524
ID AAY96524 standard: peptide: 36 AA.

XX	Key	Location/Qualifiers
FT	Modified-site	1
FT	Peptide	/note= "optionally 1
FT	Peptide	1.14
FT	Disulfide-bond	/label= TWP_1
FT		9.31
FT	Peptide	/note= "optional"
FT		15..22
FT	Peptide	/label= linker
FT		23..36
FT		/label= TWP 2

10 to 14 residues in length comprising X-2-X-1-0, X-2-X-1-1, X-2-X-1-2, X-2-X-1-4, X-1-X-1-0, X-1-X-1-1, X-1-X-1-2, X-1-X-1-3, and X-1-X-1-4. X-1 = I, A, V, L, S or R; X-2 = E, D, K or V; X-3 = G or A; X-4 = P; X-5 = T or S; X-6 = L, I, V, A or F; X-7 = R or K; X-8 = Q, N, or E; X-9 = W, Y or F; X-10 = I, I, V, A, F, M, or K; X-11 = A, L, V, L, F, S, T, K, H, or E; X-12 = A, I, V, L, F, G, S, or Q; X-13 = R, K, T, V, S, N, Q or G; X-14 = I, A, V, L, F, T, R, E, or G; L-1 = linker comprising 1 to 20 amino acids; and n = 0 or 1. The compounds bind to and activate the c-Mpl receptor which mediates the activity of endogenous thrombopoietin. The TMs are useful for increasing the production of platelets or platelet precursors (e.g. megakaryocytes) in a mammal, which is useful for treatment of diseases which involve thrombocytopenia, e.g. aplastic anaemia, immune thrombocytopenia (ITP), human immunodeficiency virus associated ITP, and systemic lupus erythematosus.

XX	Sequence	36 AA;
SO		

Query Match 83.7%; Score 159; DB 21; Length 36;
Best Local Similarity 91.7%; Pred. No. 2e-12;
Matches 33; Conservative 0; Mismatches 3; Indels 0; Gaps 0

QY 1 IEGPTLRQCLAAAGGGGGGGIEGPTLRQCLAARA 36
||||| ||||||| ||||||| ||||||| |||||||
db 1 IEGPTLRQCLAAAGGGGGGGIEGPTLRQCLAARA 36

RESULT 23
AAB17294
ID AAB17294 standard: Peptide: 37 AA:

Modified peptide; therapeutic agent; fusion; Fc domain; cancer; autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF; immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonists; MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1; cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor; endothelial growth factor; matrix metalloproteinase; thrombotic; thrombolysis; pharmaceutical

Example 1: Page 318: 608pp: English.

The present invention describes composition of matter (I) comprising an Fc domain, pharmacologically active peptides, and linkers. Where (I) is: (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2, -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4 where P1, P2, P3, and P4 = are each independently sequences of

CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 37 AA;

Query Match 82.9%; Score 157.5; DB 21; Length 37;
 Best Local Similarity 91.9%; Pred. No. 3.1e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 IEGPTLRQCLAAARA--GGGGGGGIEGPTLRQCLAAARA 36
 ||||||| ||||| ||||||| ||||||| ||||||| |||||
 Db 1 IEGPTLRQCLAAARA--GGGGGGGIEGPTLRQCLAAARA 37

RESULT 24

AAB17295
 ID AAB17295 standard; Peptide; 38 AA.

XX AAB17295;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:351.

XX Modified peptide: therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 319; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P³, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently

CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 38 AA;

Query Match 82.6%; Score 157; DB 21; Length 38;
 Best Local Similarity 89.5%; Pred. No. 3.6e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 IEGPTLRQCLAAARA--GGGGGGGIEGPTLRQCLAAARA 36
 ||||||| ||||| ||||||| ||||||| ||||||| |||||
 Db 1 IEGPTLRQCLAAARA--GGGGGGGIEGPTLRQCLAAARA 38

RESULT 25

AAB17304
 ID AAB17304 standard; Peptide; 39 AA.

XX AAB17304;

XX 31-OCT-2000 (first entry)

XX TPO-mimetic peptide sequence SEQ ID NO:360.

XX Modified peptide: therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.

XX WO200024782-A2.

XX 04-MAY-2000.

XX 25-OCT-1999; 99WO-US25044.

XX 23-OCT-1998; 98US-0105371.

XX 22-OCT-1999; 99US-0428082.

XX (AMGE-) AMGEN INC.

XX Feige U, Liu C, Cheetham J, Boone TC;

XX WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 323; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P³, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive

CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443
 CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.

XX Sequence 42 AA;
 SQ Query Match 81.6%; Score 155; DB 21; Length 42;
 Best Local Similarity 81.0%; Pred. No. 6.9e-12;
 Matches 34; Conservative 0; Mismatches 2; Indels 6; Gaps 1;
 QY 1 IEGPTLRQCLAAARA-----GGGGGGGIEGPTLRQCLAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEGPTLRQCLAAARA-----GGGGGGGIEGPTLRQCLAAARA 42

RESULT 30
 AAB17292
 ID AAB17292 standard; Peptide; 35 AA.
 XX AC AAB17292;
 XX DT 31-OCT-2000 (first entry)
 XX DE TPO-mimetic peptide sequence SEQ ID NO:348.
 XX KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
 KW autoimmune disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
 KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1, TNF; antagonist;
 KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
 KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
 KW vascular endothelial growth factor; matrix metalloproteinase;
 KW asthma; thrombosis; pharmaceutical.

XX Synthetic.
 OS WO200024782-A2.
 PN 04-MAY-2000.
 XX 25-OCT-1999; 99WO-US25044.
 XX 23-OCT-1998; 98US-0105371.
 PR 22-OCT-1999; 99US-0428082.
 XX (AMGE-) AMGEN INC.
 XX Feige U, Liu C, Cheetham J, Boone TC;
 PI WPI; 2000-350702/30.

XX Novel composition of matter comprising an Fc domain and
 PT pharmacologically active peptides, useful for treating cancer and
 PT autoimmune diseases -

XX Example 1; Page 317-318; 608pp; English.

XX The present invention describes composition of matter (I) comprising an
 CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
 CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 = are each
 CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
 CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
 CC where P1, P2, P3, and P4 = are each independently sequences of
 CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
 CC independently linkers; and a, b, c, d, e, and f = are each independently
 CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
 CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
 CC activities. DNAs, vectors and host cells from the present invention can
 CC be used for producing pharmaceutical compositions. The compositions are
 CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
 CC The use of an Fc domain (rather than a Fab domain) can provide a longer
 CC half-life or incorporate functions such as Fc receptor binding, protein
 CC A binding, complement fixation, and possibly placental transfer. AAA69443

CC to AAA69526 and AAB16955 to AAB18003 represent nucleotide and amino acid
 CC sequences used in the exemplification of the present invention.
 XX Sequence 35 AA;
 SQ Query Match 79.7%; Score 151.5; DB 21; Length 35;
 Best Local Similarity 91.7%; Pred. No. 1.5e-11;
 Matches 33; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 IEGPTLRQCLAAARAGGGGGGIEGPTLRQCLAAARA 36
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1 IEGPTLRQCLAAARA-GGGGGGIEGPTLRQCLAAARA 35

Search completed: October 9, 2002, 08:58:56
 Job time : 16.1874 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: October 9, 2002, 08:55:27 ; Search time 5.98595 Seconds
(without alignments)
146.898 Million cell updates/sec

Title: US-09-422-838c-27
Perfect score: 190
Sequence: 1 IEGLTLRQCLAAAGGGGGIEGPTLRQCLAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2.6/ptodata/2/iaa/PTIUS_COMB.pap.*
6: /cgn2.6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description
1	64.5	33.9	991	4	US-09-352-159-27
2	64.5	33.9	991	4	US-09-352-168-27
3	64.5	33.9	1196	4	US-09-352-159-31
4	64.5	33.9	1196	4	US-09-352-168-31
5	64	33.7	584	2	US-08-987-466-4
6	64	33.7	584	2	US-09-240-359-4
7	62	32.6	440	3	US-09-100-664A-2
8	62	32.6	440	3	US-09-100-664A-3
9	62	32.6	440	3	US-09-100-664A-4
10	60	31.6	14	2	US-08-764-640-13
11	60	31.6	14	2	US-08-764-640-193
12	60	31.6	14	2	US-08-973-225-13
13	60	31.6	14	2	US-08-973-225-193
14	60	31.6	14	3	US-09-244-298A-13
15	60	31.6	14	3	US-09-244-298A-193
16	60	31.6	14	4	US-09-516-704-13
17	60	31.6	14	4	US-09-516-704-193
18	60	31.6	15	2	US-08-764-640-17
19	60	31.6	15	2	US-08-764-640-185
20	60	31.6	15	3	US-08-973-225-17
21	60	31.6	15	3	US-08-973-225-185
22	60	31.6	15	3	US-09-244-298A-17
23	60	31.6	15	3	US-09-244-298A-185
24	60	31.6	15	4	US-09-516-704-17
25	60	31.6	15	4	US-09-516-704-185
26	60	31.6	16	2	US-08-764-640-18
27	60	31.6	16	2	US-08-764-640-194

28	60	31.6	16	2	US-08-764-640-232	Sequence 232, App
29	60	31.6	16	3	US-08-973-225-18	Sequence 18, Appl
30	60	31.6	16	3	US-08-973-225-194	Sequence 194, App
31	60	31.6	16	3	US-08-973-225-220	Sequence 220, App
32	60	31.6	16	3	US-09-244-298A-18	Sequence 18, Appl
33	60	31.6	16	3	US-09-244-298A-194	Sequence 194, App
34	60	31.6	16	3	US-09-244-298A-232	Sequence 232, App
35	60	31.6	16	4	US-09-516-704-18	Sequence 18, Appl
36	60	31.6	16	4	US-09-516-704-194	Sequence 194, App
37	60	31.6	16	4	US-09-516-704-232	Sequence 232, App
38	59.5	31.3	26	1	US-07-776-272-16	Sequence 16, Appl
39	59.5	31.3	126	1	US-08-451-947-57	Sequence 57, Appl
40	59.5	31.3	126	2	US-08-424-826A-57	Sequence 57, Appl
41	59.5	31.3	126	3	US-08-928-694-57	Sequence 57, Appl
42	59.5	31.3	126	5	PCT-US91-06950-57	Sequence 57, Appl
43	59.5	31.3	969	2	US-08-284-941-2	Sequence 2, Appl
44	59.5	31.3	969	2	US-08-447-642-2	Sequence 2, Appl
45	59.5	31.3	969	4	US-09-236-503-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-352-159-27
; Sequence 27, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duwick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 991
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-159-27

Query Match 33.9%; Score 64.5; DB 4; Length 991;
Best Local Similarity 54.2%; Pred. No. 5.1;
Matches 13; Conservative 3; Mismatches 7; Indels 1; Gaps 1;
QY 3 GPTLRQCL-AARAGGGGGGGIEG 25
|||: | | | | | | | | | |
Db 503 GPSIPPCADGAKAGGGGGGGSGG 526

RESULT 2
US-09-352-168-27
; Sequence 27, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duwick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875

Query Match 33.9%; Score 64.5; DB 4; Length 1196;
Best Local Similarity 54.2%; Pred. No. 6.1;
Matches 13; Conservative 3; Mismatches 7; Indels 1

Query Match 33.9%; Score 64.5; DB 4; Length 1196;
Best Local Similarity 54.2%; Pred. No. 6.1;
Matches 13: Conservative 3; Mismatches 7; Indels 1

QY 3 GPTLRQCL-AARAGGGGGGGIEG 25
|||: | | : ||||| ||| |
Db 708 GPSIPPCADGAKAGGGGGGGGG 731

RESULT 5
US-08-987-466-4
; Sequence 4, Application US/08987466
; Patent No. 5922595
; GENERAL INFORMATION:
; APPLICANT: Fisher, Douglas A.
; APPLICANT: Gooding, Doug
; APPLICANT: Streeter, Dave
; TITLE OF INVENTION: CYCLIC-GMP PHOSPHODIESTERASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

```

RESULT 5
US-08-987-466-4
; Sequence 4, Application US/08987466
; Patent No. 5922595
; GENERAL INFORMATION:
; APPLICANT: Fisher, Douglas A.
; APPLICANT: Gooding, Doug
; APPLICANT: Streeter, Dave
; TITLE OF INVENTION: CYCLIC-GMP PHOSPHODIESTERASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

```

```

:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Diskette
:
: COMPUTER: IBM Compatible
:
: OPERATING SYSTEM: DOS
:
: SOFTWARE: Fastseq for Windows Version 2.0
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/987,466
:
: FILING DATE: Filed Herewith
:
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER:
:
: FILING DATE:
:

```

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PR-0442 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PR-0442 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 829179
; US-08-987-466-4

```

Query Match 33.7%; Score 64; DB 2; Length 584;
Best Local Similarity 61.9%; Pred. No. 3.5;
Matches 13; Conservative 0; Mismatches 8; Indels

QY 11 AARAGGGGGGIEGPTLRQC 31
 | | | | | | | |
Db 555 ALRAGGGGGGGMAPRTGGC 575

RESULT 6
 US-09-240-359-4
 ; Sequence 4, Application US/09240359
 ; Patent No. 6255456
 ; GENERAL INFORMATION:
 ; APPLICANT: Fisher, Douglas A.
 ; APPLICANT: Gooding, Doug
 ; APPLICANT: Streeter, Dave
 ; TITLE OF INVENTION: CYCLIC-GMP PHOSPHODIESTERASE
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Dr.
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/240,359
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/987,466
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0442 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-855-0555
 ; TELEFAX: 650-845-4166
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 584 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GenBank
 ; CLONE: 829179
 ; US-09-240-359-4

Query Match 33.7%; Score 64; DB 4; Length 584;
 Best Local Similarity 61.9%; Pred. No. 3.5;
 Matches 13; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 11 ARAGGGGGGGGIEGPTLRQC 31
 | | | | | | | | | |
 DB 555 ALRAGGGGGGGGMAPRTGCG 575

RESULT 7
 US-09-100-664A-2
 ; Sequence 2, Application US/09100664A
 ; Patent No. 6057129
 ; GENERAL INFORMATION:
 ; APPLICANT: YOUNG, MICHAEL W.
 ; APPLICANT: KLOSS, BRIAN
 ; APPLICANT: BLAU, JUSTIN
 ; APPLICANT: PRICE, JEFFREY
 ; TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor

; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/100,664A
 ; FILING DATE: 19-JUN-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-221
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; TELEX: 133521
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 440 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; US-09-100-664A-2

Query Match 32.6%; Score 62; DB 3; Length 440;
 Best Local Similarity 55.0%; Pred. No. 4.4;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLRQCLAAARAGGGGGGGI 23
 | | | | | | | | | |
 DB 403 PERRPSIRMQGGGGGGV 422

RESULT 8
 US-09-100-664A-3
 ; Sequence 3, Application US/09100664A
 ; Patent No. 6057129
 ; GENERAL INFORMATION:
 ; APPLICANT: YOUNG, MICHAEL W.
 ; APPLICANT: KLOSS, BRIAN
 ; APPLICANT: BLAU, JUSTIN
 ; APPLICANT: PRICE, JEFFREY
 ; TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/100,664A
 ; FILING DATE: 19-JUN-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-221
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 440 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 US-09-100-664A-3

Query Match 32.6%; Score 62; DB 3; Length 440;
 Best Local Similarity 55.0%; Pred. No. 4.4;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLRQCLARAGGGGGG 23
 DB 403 PERRPSIRMRGGGGGGV 422

RESULT 9
 US-09-100-664A-4
 Sequence 4, Application US/09100664A
 Patent No. 6057129

GENERAL INFORMATION:
 APPLICANT: YOUNG, MICHAEL W.
 APPLICANT: KLOSS, BRIAN
 APPLICANT: BLAU, JUSTIN
 APPLICANT: PRICE, JEFFREY
 TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE
 TITLE OF INVENTION: THEREOF
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/100.664A
 FILING DATE: 19-JUN-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-221
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521

INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 440 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 US-09-100-664A-4

Query Match 32.6%; Score 62; DB 3; Length 440;
 Best Local Similarity 55.0%; Pred. No. 4.4;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 PTLRQCLARAGGGGGG 23

DB 403 PERRPSIRMRGGGGGGV 422

RESULT 10
 US-08-764-640-13
 Sequence 13, Application US/08764640
 Patent No. 5869451
 Patent No. 5869451 5837683
 GENERAL INFORMATION:

APPLICANT: Dower, William J.
 APPLICANT: Barrett, Ronald W.
 APPLICANT: Cwirla, Steven E.
 APPLICANT: Gates, Christian
 APPLICANT: Schatz, Peter J.
 APPLICANT: Balasubramanian, Palaniappan
 APPLICANT: Wagstrom, Christopher R.
 APPLICANT: Hendren, Richard W.
 APPLICANT: Deprience, Randolph B.
 APPLICANT: Poddaturi, Surekha
 APPLICANT: Yin, Qun
 TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
 TITLE OF INVENTION: RECEPTOR
 NUMBER OF SEQUENCES: 244
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Glaxo Wellcome
 STREET: Five Moore Drive, P.O. Box 13398
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/764,640
 FILING DATE: 11-DEC-1996
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Hrubiec, Robert T.
 REGISTRATION NUMBER: 36,392
 REFERENCE/DOCKET NUMBER: PK3281
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-248-1000
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 14 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-764-640-13

Query Match 31.6%; Score 60; DB 2; Length 14;
 Best Local Similarity 92.9%; Pred. No. 0.25;
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAARA 14
 DB 1 IEPTLRQWLAARA 14

RESULT 11
 US-08-764-640-193
 Sequence 193, Application US/08764640
 Patent No. 5869451
 Patent No. 5869451 5837683
 GENERAL INFORMATION:
 APPLICANT: Dower, William J.
 APPLICANT: Barrett, Ronald W.
 APPLICANT: Cwirla, Steven E.

APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprience, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,640
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-764-640-193
Query Match 31.6%; Score 60; DB 2; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 IEPTLRQCLAARA 14
Db 1 IEPTLRQWLAARA 14
RESULT 12
US-08-973-225-13
Sequence 13, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW

STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-08-973-225-13
Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 IEPTLRQCLAARA 14
Db 1 IEPTLRQWLAARA 14
RESULT 13
US-08-973-225-193
Sequence 193, Application US/08973225A
Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Mattheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-244-298A-13

Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAARA 14
|||||
Db 1 IEGPTLRQWLAARA 14

RESULT 14
US-09-244-298A-13
; Sequence 13, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-244-298A-13

Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;

Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 IEGPTLRQCLAARA 14
|||||
Db 1 IEGPTLRQWLAARA 14

RESULT 15
US-09-244-298A-193
; Sequence 193, Application US/09244298A
; Patent No. 6121238
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirila, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/244,298A
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-244-298A-193

Query Match 31.6%; Score 60; DB 3; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAARA 14
|||||
Db 1 IEGPTLRQWLAARA 14

RESULT 16
US-09-516-704-13
; Sequence 13, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.

;
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-516-704-13

Query Match 31.6%; Score 60; DB 4; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAAARA 14

RESULT 17
US-09-516-704-193
; Sequence 193, Application US/09516704
; Patent No. 6251864
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Gates, Christian
; Schatz, Peter J.
; Balasubramanian, Palaniappan
; Wagstrom, Christopher R.
; Hendren, Richard W.
; Deprince, Randolph B.
; Podduturi, Surekha
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398

;
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,704
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-516-704-193

Query Match 31.6%; Score 60; DB 4; Length 14;
Best Local Similarity 92.9%; Pred. No. 0.25;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCLAAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPTLRQWLAAARA 14

RESULT 18
US-08-764-640-17
; Sequence 17, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996

```
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-764-640-17

Query Match 31.6%; Score 60; DB 2; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGLPTLRQCLAARA 14
   ||||| |||||
Db 1 IEGLPTLRQWLAARA 14

RESULT 19
US-08-764-640-185
; Sequence 185, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: Deprince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 244
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

; MOLECULE TYPE: peptide
US-08-764-640-185

Query Match 31.6%; Score 60; DB 2; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGLPTLRQCLAARA 14
   ||||| |||||
Db 1 IEGLPTLRQWLAARA 14

RESULT 20
US-08-973-225-17
; Sequence 17, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwiria, Steven E.
; APPLICANT: Duffin, David J.
; APPLICANT: Gates, Christian
; APPLICANT: Haseiden, Sherrill S.
; APPLICANT: Matheakis, Larry C.
; APPLICANT: Schatz, Peter J.
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Wrighton, Nicholas C.
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK306505W
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-08-973-225-17

Query Match 31.6%; Score 60; DB 3; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGLPTLRQCLAARA 14
   ||||| |||||
Db 1 IEGLPTLRQWLAARA 14

RESULT 21
US-08-973-225-185
; Sequence 185, Application US/08973225A
```


Patent No. 6083913
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwiria, Steven E.
Duffin, David J.
Gates, Christian
Haselden, Sherril S.
Matheakis, Larry C.
Schatz, Peter J.
Wagstrom, Christopher R.
Wrighton, Nicholas C.
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
THROMBOPOIETIN RECEPTOR
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,225A
FILING DATE: 04-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3065USW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-08-973-225-185
Query Match 31.6%; Score 60; DB 3; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27; 1; Indels 0; Gaps 0;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 IEPTLRQCLAA 14
| | | | | | | | | | | | | | | |
Db 2 IEPTLRQCLAA 15
| | | | | | | | | | | | | | | |
RESULT 22
US-09-244-298A-17
Sequence 17, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-17
Query Match 31.6%; Score 60; DB 3; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27; 1; Indels 0; Gaps 0;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 IEPTLRQCLAA 14
| | | | | | | | | | | | | | | |
Db 1 IEPTLRQCLAA 14
| | | | | | | | | | | | | | | |
RESULT 23
US-09-244-298A-185
Sequence 185, Application US/09244298A
Patent No. 6121238
GENERAL INFORMATION:
APPLICANT: Dower, William J.
APPLICANT: Barrett, Ronald W.
APPLICANT: Cwiria, Steven E.
APPLICANT: Gates, Christian
APPLICANT: Schatz, Peter J.
APPLICANT: Balasubramanian, Palaniappan
APPLICANT: Wagstrom, Christopher R.
APPLICANT: Hendren, Richard W.
APPLICANT: Deprence, Randolph B.
APPLICANT: Podduturi, Surekha
APPLICANT: Yin, Qun
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
RECEPTOR
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/244,298A
FILING DATE: 11-DEC-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-244-298A-185

Query Match 31.6%; Score 60; DB 3; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPLTROCCLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPLTROWLAARA 15

RESULT 24
US-09-516-704-17
Sequence 17, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-516-704-17

Query Match 31.6%; Score 60; DB 4; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPLTROCCLAARA 14
| | | | | | | | | | | | | | | |
Db 1 IEGPLTROWLAARA 14

RESULT 25
US-09-516-704-185
Sequence 185, Application US/09516704
Patent No. 6251864
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Barrett, Ronald W.
Cwirla, Steven E.
Gates, Christian
Schatz, Peter J.
Balasubramanian, Palaniappan
Wagstrom, Christopher R.
Hendren, Richard W.
Deprince, Randolph B.
Podduturi, Surekha
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
NUMBER OF SEQUENCES: 244
CORRESPONDENCE ADDRESS:
ADDRESSEE: Glaxo Wellcome
STREET: Five Moore Drive, P.O. Box 13398
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 01-Mar-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hrubiec, Robert T.
REGISTRATION NUMBER: 36,392
REFERENCE/DOCKET NUMBER: PK3281
TELEPHONE: 919-248-1000
INFORMATION FOR SEQ ID NO: 185:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 185:
US-09-516-704-185

Query Match 31.6%; Score 60; DB 4; Length 15;
Best Local Similarity 92.9%; Pred. No. 0.27;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPLTROCCLAARA 14
| | | | | | | | | | | | | | | |
Db 2 IEGPLTROWLAARA 15

RESULT 26

US-08-764-640-18
; Sequence 18, Application US/08764640
; Patent No. 5869451
; Patent No. 5869451 5837683
; GENERAL INFORMATION:

; APPLICANT: Dower, William J.
; APPLICANT: Barrett, Ronald W.
; APPLICANT: Cwirla, Steven E.
; APPLICANT: Gates, Christian
; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: DepPrince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun

; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

; TITLE OF INVENTION: RECEPTOR

; NUMBER OF SEQUENCES: 244

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park

; STATE: NC

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/764,640

; FILING DATE: 11-DEC-1996

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Hrubiec, Robert T.

; REGISTRATION NUMBER: 36,392

; REFERENCE/DOCKET NUMBER: PK3281

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-248-1000

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 16 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 15

; OTHER INFORMATION: /product= "Beta-ala"

US-08-764-640-18

Query Match 31.6%; Score 60; DB 2; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAARA 14

Db 1 IEPTLRQWLAARA 14

RESULT 27

US-08-764-640-194

; Sequence 194, Application US/08764640

; Patent No. 5869451

; Patent No. 5869451 5837683

; GENERAL INFORMATION:

; APPLICANT: Dower, William J.

; APPLICANT: Barrett, Ronald W.

; APPLICANT: Cwirla, Steven E.

; APPLICANT: Gates, Christian

; APPLICANT: Schatz, Peter J.
; APPLICANT: Balasubramanian, Palaniappan
; APPLICANT: Wagstrom, Christopher R.
; APPLICANT: Hendren, Richard W.
; APPLICANT: DepPrince, Randolph B.
; APPLICANT: Podduturi, Surekha
; APPLICANT: Yin, Qun

; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

; TITLE OF INVENTION: RECEPTOR

; NUMBER OF SEQUENCES: 244

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Glaxo Wellcome

; STREET: Five Moore Drive, P.O. Box 13398

; CITY: Research Triangle Park

; STATE: NC

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/764,640

; FILING DATE: 11-DEC-1996

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Hrubiec, Robert T.

; REGISTRATION NUMBER: 36,392

; REFERENCE/DOCKET NUMBER: PK3281

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-248-1000

; INFORMATION FOR SEQ ID NO: 194:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 16 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-764-640-194

Query Match 31.6%; Score 60; DB 2; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEPTLRQCLAARA 14

Db 2 IEPTLRQWLAARA 15

RESULT 28

US-08-764-640-232

; Sequence 232, Application US/08764640

; Patent No. 5869451

; Patent No. 5869451 5837683

; GENERAL INFORMATION:

; APPLICANT: Dower, William J.

; APPLICANT: Barrett, Ronald W.

; APPLICANT: Cwirla, Steven E.

; APPLICANT: Gates, Christian

; APPLICANT: Schatz, Peter J.

; APPLICANT: Balasubramanian, Palaniappan

; APPLICANT: Wagstrom, Christopher R.

; APPLICANT: Hendren, Richard W.

; APPLICANT: DepPrince, Randolph B.

; APPLICANT: Podduturi, Surekha

; APPLICANT: Yin, Qun

; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A

; TITLE OF INVENTION: RECEPTOR

; NUMBER OF SEQUENCES: 244

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Glaxo Wellcome

; STREET: Five Moore Drive, P.O. Box 13398

```
;
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-232
;
; Query Match 31.6%; Score 60; DB 2; Length 16;
; Best Local Similarity 92.9%; Pred. No. 0.29;
; Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQCLAARA 14
; Db 2 IEGPTLRQWLAARA 15
;
; RESULT 29
; US-08-973-225-18
; Sequence 18, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
;
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,640
; FILING DATE: 11-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 232:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-764-640-232
;
; Query Match 31.6%; Score 60; DB 2; Length 16;
; Best Local Similarity 92.9%; Pred. No. 0.29;
; Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 IEGPTLRQCLAARA 14
; Db 2 IEGPTLRQWLAARA 15
;
; RESULT 30
; US-08-973-225-194
; Sequence 194, Application US/08973225A
; Patent No. 6083913
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Barrett, Ronald W.
; Cwirla, Steven E.
; Duffin, David J.
; Gates, Christian
; Haselden, Sherril S.
; Mattheakis, Larry C.
; Schatz, Peter J.
; Wagstrom, Christopher R.
; Wrighton, Nicholas C.
;
; TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glaxo Wellcome
; STREET: Five Moore Drive, P.O. Box 13398
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,225A
; FILING DATE: 04-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hrubiec, Robert T.
; REGISTRATION NUMBER: 36,392
; REFERENCE/DOCKET NUMBER: PK3065USW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-248-1000
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 194:
```

US-08-973-225-194

Query Match 31.6%; Score 60; DB 3; Length 16;
Best Local Similarity 92.9%; Pred. No. 0.29;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 IEGPTLRQCIARA 14
 |||||||
Db 2 IEGPTLRQWLAARA 15

Search completed: October 9, 2002, 09:06:31
Job time : 6.98595 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:54:17 ; Search time 8.09368 seconds
(without alignments)
427.397 Million cell updates/sec

Title: US-09-422-838c-27
Perfect score: 190
Sequence: 1 IEPTLRQCLAAAGGGGGGIEPTLRQCLAAARA 36

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 28138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 28138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	38.4	619	1 KSNCL	laccase (EC 1.10.3
2	73	38.4	619	1 KSNCL	laccase (EC 1.10.3
3	68	35.8	199	2 T48099	hypothetical prote
4	66	34.7	500	2 T20961	hypothetical prote
5	65.5	34.5	403	2 A53662	homeotic protein H
6	65	34.2	77	1 INSH	insulin precursor
7	65	34.2	105	1 IPBO	insulin precursor
8	65	34.2	201	2 J01094	hypothetical 20.2K
9	64.5	33.9	1733	1 B45344	probable nuclear a
10	64	33.7	331	2 T26807	hypothetical prote
11	64	33.7	333	2 T26808	hypothetical prote
12	64	33.7	777	2 S65543	3',5'-cyclic-nucle
13	63.5	33.4	434	2 T47772	hypothetical prote
14	63	33.2	488	2 G87033	probable ATP/Grp-b
15	63	33.2	518	2 S72938	hflx protein - Myc
16	63	33.2	806	2 T13690	hypothetical prote
17	63	33.2	1168	1 MWAXIC	myosin heavy chain
18	62.5	32.9	339	2 T06612	hypothetical prote
19	62	32.6	201	2 T49792	hypothetical prote
20	62	32.6	867	2 S57795	probable deoxyribo
21	62	32.6	889	2 T09055	protocadherin 68 -
22	61	32.1	495	2 D70505	probable Hflx - My
23	60	31.6	167	2 S71779	glycine-rich RNA-b
24	60	31.6	285	2 S69312	probable membrane
25	60	31.6	323	2 S20099	transforming prote
26	60	31.6	649	2 S58084	hdc protein - frui
27	60	31.6	1325	2 T13386	hypothetical prote
28	59.5	31.3	443	2 E96495	hypothetical prote
29	59.5	31.3	487	2 B39490	subtilisin-like, pr

30 59.5 31.3 652 1 JC2191 subtilisin-like pr
31 59.5 31.3 962 2 JC5571 subtilisin-like pr
32 59.5 31.3 969 1 A39490 subtilisin-like pr
33 59.5 31.3 975 2 JC5570 subtilisin-like pr
34 59 31.1 102 2 H95992 hypothetical prote
35 59 31.1 165 2 S41773 glycine-rich RNA-b
36 59 31.1 165 2 S59529 RNA-binding glycin
37 59 31.1 250 2 H85067 hypothetrical prote
38 59 31.1 298 2 C96690 unknown protein F2
39 59 31.1 346 1 S35500 heterogeneous ribo
40 59 31.1 367 2 JC6087 helix-loop-helix t
41 59 31.1 396 2 T49109 glycine-rich prote
42 59 31.1 517 2 B71260 hypothetrical prote
43 59 31.1 543 2 F96624 hypothetrical prote
44 59 31.1 593 1 KRHU0 keratin 10, type I
45 59 31.1 1428 2 T13926 probable protein p

ALIGNMENTS

RESULT 1
KSNCL
laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain OR)
N;Alternate names: urishiol oxidase
C;Species: Neurospora crassa
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
C;Accession: A28523; A29762
R;Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
J. Biol. Chem. 263, 885-896, 1988
A;Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- a
A;Reference number: A28523; MUID:88087214
A;Accession: A28523
A;Molecule type: DNA
A;Residues: 1-619 <GER>
A;Cross-references: EMBL:M14554
R;Germann, U.A.; Lerch, K.
Proc. Natl. Acad. Sci. U.S.A. 83, 8854-8858, 1986
A;Title: Isolation and partial nucleotide sequence of the laccase gene from Neurospor
A;Reference number: A29762; MUID:87067412
A;Accession: A29762
A;Molecule type: DNA
A;Residues: 379-619 <GE2>
A;Cross-references: GB:M14554; NID:gl68823; PIDN:AAA33590.1; PID:gl68824
C;Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquino
C;Genetics:
A;Introns: 86/3
C;Superfamily: laccase
C;Keywords: copper; glycoprotein; oxidoreductase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-49/Domain: propeptide #status predicted <PRO>
F;50-619/Product: laccase #status predicted <MAT>
F;84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
F;216-372/Domain: middle beta-barrel #status predicted <BB2>
F;431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
F;139,282,295,340,422,444/Binding site: carbohydrate (Asn) (covalent) #status predict
F;144,480/Binding site: copper (His) (type 2) #status predicted
F;146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status p
F;477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 38.4%; Score 73; DB 1; Length 619;
Best Local Similarity 60.0%; Pred. No. 1.2;
Matches 15; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 7 RQCLAAAGGGGGGIEPTLRQC 31
||| ||||| |||||
Db 39 RQDSQAERYGGGGGNCSPNRC 63

RESULT 2
KSNCL
laccase (EC 1.10.3.2) precursor - Neurospora crassa (strain TS)
N;Alternate names: urishiol oxidase

C:Species: Neurospora crassa
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 11-Jun-1999
 C:Accession: B28523
 R:Germann, U.A.; Mueller, G.; Hunziker, P.E.; Lerch, K.
 J. Biol. Chem. 263, 885-896, 1988
 A:Title: Characterization of two allelic forms of Neurospora crassa laccase. Amino- and
 A:Reference number: A28523; MUID:88087214
 A:Accession: B28523
 A:Molecule type: DNA
 A:Residues: 1-619 <GER>
 A:Cross-references: EMBL:M18334; NID:gl68827; PIDN:AAA33592.1; PID:gl68828
 C:Comment: This enzyme, which catalyzes the oxidation of benzendiol to benzosemiquinone
 C:Genetics:
 A:Introns: 86/3
 C:Superfamily: laccase
 C:Keywords: copper; glycoprotein; oxidoreductase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-49/Domain: propeptide #status predicted <PRO>
 F:50-619/Product: laccase #status predicted <MAT>
 F:84-215/Domain: amino-terminal beta-barrel #status predicted <BB1>
 F:216-372/Domain: middle beta-barrel #status predicted <BB2>
 F:431-580/Domain: carboxyl-terminal beta-barrel #status predicted <BB3>
 F:139,282,295,340,422,444/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:144,480/Binding site: copper (His) (type 2) #status predicted
 F:146,189,191,482,548,550/Binding site: 2Cu-O cluster (His) (copper type 3) #status predicted
 F:477,549,554/Binding site: copper (His, Cys, His) (type 1) #status predicted

Query Match 38.4%; Score 73; DB 1; Length 619;
 Best Local Similarity 60.0%; Pred. No. 1.2;
 Matches 15; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 7 ROCLEARAGGGGGGIEGPTLRQC 31
 || | ||||| |||||
 Db 39 RQDSQERYGGGGGCSPTNRQC 63

RESULT 3
 T48099
 hypothetical protein T20010.200 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T48099
 R:Obermaier, B.; Ottenwaelder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Rudd, S.; Lem
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z24484
 A:Accession: T48099
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-199 <OBE>
 A:Cross-references: EMBL:AL163816
 A:Experimental source: cultivar Columbia; BAC clone T20010
 C:Genetics:
 A:Map position: 3
 A:Introns: 163/2
 A:Note: T20010.200

Query Match 35.8%; Score 68; DB 2; Length 199;
 Best Local Similarity 34.8%; Pred. No. 1.5;
 Matches 16; Conservative 3; Mismatches 11; Indels 16; Gaps 1;

QY 2 EGPTLRQC-----LAARAGGGGGGIEGPTLRQC 31
 || | | | |||||:| | |
 Db 7 EGRTRCPASTTCSTLVAQTSLLCVDDGGGGGVDGVDRC 52

RESULT 4
 T20961
 hypothetical protein F15B9.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20961
 R:Percy, C.
 submitted to the EMBL Data Library, August 1996.

A:Reference number: Z19351
 A:Accession: T20961
 A:Status: preliminary; translated from GB/EMBL/DBBJ
 A:Molecule type: DNA
 A:Residues: 1-500 <WIL>
 A:Cross-references: EMBL:Z78013; PIDN:CA01420.1; GSPDB:GN00023; CESP:F15B9.5
 A:Experimental source: clone F15B9
 C:Genetics:
 A:Gene: CESP:F15B9.5
 A:Map position: 5
 A:Introns: 46/3; 63/3; 125/2; 162/2; 283/3; 391/1; 446/1

Query Match 34.7%; Score 66; DB 2; Length 500;
 Best Local Similarity 56.5%; Pred. No. 5.5;
 Matches 13; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 GPTLRQCLAAAGGGGGGIEG 25
 | | | | | |||||:| |
 Db 429 GMLGRFLSNRGGGGGGGGG 451

RESULT 5
 A53662
 homeotic protein HB9 - human
 C:Species: Homo sapiens (man)
 C:Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 17-Oct-1997
 C:Accession: A53662
 R:Harrison, K.A.; Druet, K.M.; Deguchi, Y.; Tusciano, J.M.; Kehrl, J.H.
 J. Biol. Chem. 269, 19668-19975, 1994
 A:Title: A novel human homeobox gene distantly related to proboscipedia is expressed
 A:Reference number: A53662; MUID:94327547
 A:Accession: A53662
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-403 <HAR>
 A:Cross-references: GB:U07663
 A:Note: the nucleotide sequence and conceptual translation as given are self-consiste
 C:Genetics:
 A:Gene: GDB:HLXB9
 A:Cross-references: GDB:I36411; OMIM:142994
 A:Map position: 1q41-1q42.1
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:244-300/Domain: homeobox homology <HAX>

Query Match 34.5%; Score 65.5; DB 2; Length 403;
 Best Local Similarity 57.1%; Pred. No. 5.2;
 Matches 16; Conservative 0; Mismatches 9; Indels 3; Gaps 1;

QY 10 LAARA---GGGGGGGIEGPTLRQCLAA 34
 || | | ||||| | | | |
 Db 34 LAAAAAGTGGGGGGGASGTTGSCSPA 61

RESULT 6
 INSH
 insulin precursor - sheep
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
 C:Accession: S16430; S16431
 R:Brown, H.; Sanger, F.; Kitai, R.
 Biochem. J. 60, 556-565, 1955
 A:Title: The structure of pig and sheep insulins.
 A:Reference number: A90344
 A:Accession: S16430
 A:Molecule type: protein
 A:Residues: 1-30:57-77 <BRO>
 R:Peterson, J.D.; Nehrlich, S.; Over, P.E.; Steiner, D.F.
 J. Biol. Chem. 247, 4866-4871, 1972
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proin
 A:Reference number: A92111; MUID:72258016
 A:Accession: S16431
 A:Molecule type: protein

A;Molecule type: protein
A;Residues: 85-105 <CHE>
A;Accession: S48185
A;Status: preliminary
A;Molecule type: protein
A;Residues: 25-30,'X',32-42,'X',44-54 <CH2>
R;Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.
Biochem. J. 60, 541-556, 1955
A;Title: The disulphide bonds of insulin.
A;Reference number: A90343
A;Contents: annotation; amides; disulfides
R;Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Baumeister, W.
FEBS Lett. 349, 205-209, 1994
A;Title: Existence of a molecular ruler in proteasomes suggested by analysis of degradation products
A;Reference number: S46258; MUID:94326921
A;Accession: S46258
A;Status: preliminary
A;Molecule type: protein
A;Residues: 25-54 <WEN>
C;Superfamily: Insulin
C;Keywords: hormone; pancreas
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54/Domain: insulin chain B #status experimental <BCH>
F;25-54,85-105/Product: insulin #status experimental <NAT>
F;57-82/Domain: connecting peptide #status experimental <CPPE>
F;85-105/Domain: insulin chain A #status experimental <ACH>
F;31-91,43-104,90-95/Disulfide bonds: #status experimental

Query Match 34.2%; Score 65; DB 1; Length 105;
Best Local Similarity 41.0%; Pred. No. 1.8;
Matches 16; Conservative 4; Mismatches 11; Indels 8; Gaps 2;

QY 1 IEPTLRQCLARAGGGGGGTEGP-----TLRQCILAA 34
| | | | | | | | | | | | | | : | | |
Db 58 VEGP---QVGALELAGPGAGGLEPPKRGIVEQQCAS 93

RESULT 8
JQ1094
hypothetical 20.2K protein - tomato ringspot virus
C;Species: tomato ringspot virus
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 08-Oct-1999
C;Accession: JQ1094
R;Rott, M.E.; Tremaine, J.H.; Rochon, D.M.
J. Gen. Virol. 72, 1505-1514, 1991
A;Title: Nucleotide sequence of tomato ringspot virus RNA-2.
A;Reference number: JQ1093; MUID:91311402
A;Accession: JQ1094
A;Status: translation not shown
A;Molecule type: genomic RNA
A;Residues: 1-201 <ROT>
A;Cross-references: GB:D12477; GB:D01129; MID:g222674; PIDN:BARO2044.1; PID:d1002526;
A;Experimental source: strain raspberry

Query Match 34.2%; Score 65; DB 2; Length 201;
Best Local Similarity 61.5%; Pred. No. 3.2;
Matches 16; Conservative 1; Mismatches 5; Indels 4; Gaps 1;

QY 13 RAGGGGGGGGIE---GPTLRQCILAA 34
| | | | | | | | | | | | | | | | | |
Db 13 RAGGGGGGGGKEVFKAgrTLTKVLKA 38

RESULT 9
B45344
probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C;Species: suid herpesvirus 1
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: B45344
R;Vicek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzler, M.
Virology 179, 365-377, 1990
A;Title: pseudorabies virus immediate-early gene overlaps with an oppositely oriented
A;Reference number: A45344; MUID:91021039

C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: Tl3690
R;Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: 217699
A;Accession: Tl3690
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-806 <MUR>
A;Cross-references: EMBL:AL031863; NID:el331652; PID:el355938; PIDN:CAA21318.
C;Genetics:
A;Cross-references: FlyBase:FBgn025833
A;Introns: 37/3; 448/3; 611/2; 690/3
A;Note: EG:EG0003.2

Query Match 33.2%; Score 63; DB 2; Length 806;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps

QY 15 GGGGGGEGPTLRQCLAA 36
||||||| : | : |
DB 100 GGGGGGPGASITQIAQA 121
||||||| : | : |

RESULT 17
MWAXIC
myosin heavy chain IC - Acanthamoeba castellanii
N;Contains: myosin ATPase (EC 3.6.1.32)
C;Species: Acanthamoeba castellanii
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
A;Accession: A33891; C34448; A24146
R;Jung, G.; Korn, E.D.; Hammer III, J.A.
Proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987
A;Title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
A;Reference number: A33891; MUID:88016163
A;Accession: A33891
A;Molecule type: DNA
A;Residues: 1-1168 <JUN>
A;Cross-references: GB:J02974; NID:gl55624; PIDN:AAA27707.1; PID:gl55625
A;Note: this gene and protein are called MIB in this paper
R;Brzeska, H.; Lynch, T.J.; Martin, B.; Korn, E.D.
J. Biol. Chem. 264, 19340-19348, 1989
A;Title: The localization and sequence of the phosphorylation sites of Acant
A;Reference number: A34448; MUID:90037074
A;Accession: C34448
A;Molecule type: protein
A;Residues: 308-314,'X',316-329 <BRZ>
A;Comment: In this protein, the coiled-coil rod-like region found in many myo
he protein is globular and does not self-associate into filaments.
C;Genetics:
A;Gene: MIC
A;Introns: 1/3; 37/3; 60/2; 100/2; 153/3; 179/3; 208/2; 242/3; 287/3; 321/3;
A;Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology
C;Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosph
F;10-653/Domain: myosin motor domain homology <MMOT>
F;101-108/Region: nucleotide-binding motif A (P-loop)
F;543-564/Region: actin binding #status predicted
F;671-1168/Domain: carboxyl-terminal <CTD>
F;675-883/Region: basic
F;923-978/Region: alanine/glycine/proline-rich
F;983-1030/Domain: SH3 homology <SH3>
F;1034-1168/Region: alanine/glycine/proline-rich
F;107/Binding site: ATP (Lys) #status predicted
F;311/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 33.2%; Score 63; DB 1; Length 1168;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps

QY 8 OCLAAAGGGGGGIEGPT 27
| : : ||||| | : |
DB 920 QILKANGGGGGGRGRGPS 939
| : : ||||| | : |

C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: Tl3690
R;Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: 217699
A;Accession: Tl3690
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-806 <MUR>
A;Cross-references: EMBL:AL031863; NID:el331652; PID:el355938; PIDN:CAA21318.
C;Genetics:
A;Cross-references: FlyBase:FBgn025833
A;Introns: 37/3; 448/3; 611/2; 690/3
A;Note: EG:EG0003.2

Query Match 33.2%; Score 63; DB 2; Length 806;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps

QY 15 GGGGGGEGPTLRQCLAA 36
||||||| : | : |
DB 100 GGGGGGPGASITQIAQA 121
||||||| : | : |

RESULT 17
MWAXIC
myosin heavy chain IC - Acanthamoeba castellanii
N;Contains: myosin ATPase (EC 3.6.1.32)
C;Species: Acanthamoeba castellanii
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
A;Accession: A33891; C34448; A24146
R;Jung, G.; Korn, E.D.; Hammer III, J.A.
Proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987
A;Title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
A;Reference number: A33891; MUID:88016163
A;Accession: A33891
A;Molecule type: DNA
A;Residues: 1-1168 <JUN>
A;Cross-references: GB:J02974; NID:gl55624; PIDN:AAA27707.1; PID:gl55625
A;Note: this gene and protein are called MIB in this paper
R;Brzeska, H.; Lynch, T.J.; Martin, B.; Korn, E.D.
J. Biol. Chem. 264, 19340-19348, 1989
A;Title: The localization and sequence of the phosphorylation sites of Acant
A;Reference number: A34448; MUID:90037074
A;Accession: C34448
A;Molecule type: protein
A;Residues: 308-314,'X',316-329 <BRZ>
A;Comment: In this protein, the coiled-coil rod-like region found in many myo
he protein is globular and does not self-associate into filaments.
C;Genetics:
A;Gene: MIC
A;Introns: 1/3; 37/3; 60/2; 100/2; 153/3; 179/3; 208/2; 242/3; 287/3; 321/3;
A;Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology
C;Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosph
F;10-653/Domain: myosin motor domain homology <MMOT>
F;101-108/Region: nucleotide-binding motif A (P-loop)
F;543-564/Region: actin binding #status predicted
F;671-1168/Domain: carboxyl-terminal <CTD>
F;675-883/Region: basic
F;923-978/Region: alanine/glycine/proline-rich
F;983-1030/Domain: SH3 homology <SH3>
F;1034-1168/Region: alanine/glycine/proline-rich
F;107/Binding site: ATP (Lys) #status predicted
F;311/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 33.2%; Score 63; DB 1; Length 1168;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps

QY 8 OCLAAAGGGGGGIEGPT 27
| : : ||||| | : |
DB 920 QILKANGGGGGGRGRGPS 939
| : : ||||| | : |

C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: Tl3690
R;Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: 217699
A;Accession: Tl3690
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-806 <MUR>
A;Cross-references: EMBL:AL031863; NID:el331652; PID:el355938; PIDN:CAA21318.
C;Genetics:
A;Cross-references: FlyBase:FBgn025833
A;Introns: 37/3; 448/3; 611/2; 690/3
A;Note: EG:EG0003.2

Query Match 33.2%; Score 63; DB 2; Length 806;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 12; Conservative 3; Mismatches 7; Indels 0; Gaps

QY 15 GGGGGGEGPTLRQCLAA 36
||||||| : | : |
DB 100 GGGGGGPGASITQIAQA 121
||||||| : | : |

RESULT 17
MWAXIC
myosin heavy chain IC - Acanthamoeba castellanii
N;Contains: myosin ATPase (EC 3.6.1.32)
C;Species: Acanthamoeba castellanii
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 19-Jan-2001
A;Accession: A33891; C34448; A24146
R;Jung, G.; Korn, E.D.; Hammer III, J.A.
Proc. Natl. Acad. Sci. U.S.A. 84, 6720-6724, 1987
A;Title: The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like
A;Reference number: A33891; MUID:88016163
A;Accession: A33891
A;Molecule type: DNA
A;Residues: 1-1168 <JUN>
A;Cross-references: GB:J02974; NID:gl55624; PIDN:AAA27707.1; PID:gl55625
A;Note: this gene and protein are called MIB in this paper
R;Brzeska, H.; Lynch, T.J.; Martin, B.; Korn, E.D.
J. Biol. Chem. 264, 19340-19348, 1989
A;Title: The localization and sequence of the phosphorylation sites of Acant
A;Reference number: A34448; MUID:90037074
A;Accession: C34448
A;Molecule type: protein

RESULT 18
T06612
hypothetical protein F16J13.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C:Accession: T06612
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15789
A:Accession: T06612
A:Molecule type: DNA
A:Residues: 1-339 <BEV>
A:Cross-references: EMBL:AL049638; GSPDB:GN00062; ATSP:F16J13.120
A:Experimental source: cultivar Columbia; BAC clone F16J13
C:Genetics:
A:Gene: ATSP:F16J13.120
A:Map position: 4
C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 32.9%; Score 62.5; DB 2; Length 339;
Best Local Similarity 52.0%; Pred. No. 9.3;
Matches 13; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 15 GCGGGGGGTEGPTL---RQCLAARA 36
||||||| : : : : :
Db 282 GCGGGGGGSPPMGQQQAMAAMA 306

RESULT 19
T49792
hypothetical protein B9J10.290 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49792
R:Schulte, U.; Aign, V.; Hohseil, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49792
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <SCH>
A:Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSP:B9J10.290
A:Experimental source: BAC clone B9J10; strain OR74A
C:Genetics:
A:Gene: NCSP:B9J10.290
A:Map position: 6

Query Match 32.6%; Score 62; DB 2; Length 201;
Best Local Similarity 76.9%; Pred. No. 6.7;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 13 RAGGGGGGGGIEG 25
||||||| :
Db 74 RGGGGGGGGVNG 86

RESULT 20
S57795
probable deoxyribodiprimidine photo-lyase (EC 4.1.99.3) - Chlamydomonas reinhardtii
N:Alternate names: DNA photolyase homolog; probable blue light photoreceptor
C:Species: Chlamydomonas reinhardtii
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57795; S66368
R:Small, G.D.; Mip, B.; Lefebvre, P.A. Plant Mol. Biol. 28, 443-454, 1995
A:Title: Characterization of a Chlamydomonas reinhardtii gene encoding a protein of the
A:Reference number: S57795; MUID:95359403
A:Accession: S57795
A:Molecule type: DNA
A:Residues: 1-867 <SWA>
A:Cross-references: EMBL:L07561; NID:9945420; PIDN:AAC37438.1; PID:9945421

A:Accession: S66368
A:Molecule type: mRNA
A:Residues: 1-867 <SMW>
A:Cross-references: EMBL:L07561; NID:9945420; PIDN:AAC37438.1; PID:9945421
C:Genetics:
A:Gene: CPH1
C:Keywords: carbon-carbon lyase; photoreceptor

Query Match 32.6%; Score 62; DB 2; Length 867;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 4 PTLRQCLAARAGGGGGGIEG 25
| : | ||||| :
Db 713 PQMLDAAAGAGGGGGGGLAG 734

RESULT 21
T09055
protocadherin 68 - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T09055
R:Jin, P.; Xu, H.; Israel, D. submitted to the EMBL Data Library, October 1997
A:Reference number: Z16540
A:Accession: T09055
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-889 <JIN>
A:Cross-references: EMBL:AF029343; NID:92599501; PID:g2599502
C:Genetics:
A:Gene: PCH68

Query Match 32.6%; Score 62; DB 2; Length 889;
Best Local Similarity 57.9%; Pred. No. 24;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 8 QCLAARAGGGGGGIEGP 26
| : | ||||| :
Db 388 QCRVLGGGTGGGGGLGGP 406

RESULT 22
D70505
probable HflX - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 02-Sep-2000
C:Accession: D70505
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G. A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: D70505
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-495 <COL>
A:Cross-references: GB:Z98209; GB:AL123456; NID:g3261838; PIDN:CAB10901.1; PID:e33228
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: hflX
C:Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match 32.1%; Score 61; DB 2; Length 495;
Best Local Similarity 45.8%; Pred. No. 19;
Matches 11; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 3 GPTLRQCLAARAGGGGGGIEGP 26
| : : | ||||| : :
| : : | ||||| : : |

Wed Oct 9 10:29:52 2002

Db 205 GESMSRQAGGAGGGGVLGRP 228

RESULT 23

S71779

C:Species: Triticum aestivum (common wheat)

C:Date: 04-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 23-Jul-1999

C:Accession: S71779

R:Gullinan, M.J.; Niu, X.

Plant Mol. Biol. 30, 1301-1306, 1996

A:Title: cDNA encoding a wheat (Triticum aestivum cv. Chinese spring) glycine-rich RNA-binding protein

A:Reference number: S71779; MUID:96311016

A:Accession: S71779

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-167 <GUI>

A:Cross-references: EMBL:U32310; NID:g974604; PIDN:AAA75104.1; PID:g974605

C:Genetics:

A:Gene: GRP1

C:Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology

F:7-74/Domain: ribonucleoprotein repeat homology <RRM4>

Query Match 31.6%; Score 60; DB 2; Length 167;

Best Local Similarity 55.6%; Pred. No. 9.3; Indels 0; Gaps 0;

Matches 10; Conservative 4; Mismatches 4;

QY 5 TLROCLAARAGGGGGGG 22

1 : : 1-11111111

Db 78 TVNEAQRSGGGGGGG 95

RESULT 24

S69312

probable membrane protein YLR338w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein L8300.13-a

C:Species: Saccharomyces cerevisiae

C:Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 05-Nov-1999

C:Accession: S69312

R:Du, Z.

submitted to the EMBL Data Library, January 1994

A:Description: The sequence of S. cerevisiae cosmid 8300.

A:Reference number: S69312

A:Accession: S69312

A:Molecule type: DNA

A:Residues: 1-285 <DUZ>

A:Cross-references: EMBL:U19028; NID:g609380; PID:g2340034; GSPDB:GN00012; MIPS:YLR338w

C:Genetics:

A:Gene: MIPS:YLR338w

A:Map position: 12R

C:Keywords: transmembrane protein

F:142-158/Domain: transmembrane #status predicted <TM1>

F:201-217/Domain: transmembrane #status predicted <TM2>

Query Match 31.6%; Score 60; DB 2; Length 285;

Best Local Similarity 57.9%; Pred. No. 15;

Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 10 LAARAGGGGGGIEGPTL 28

1 : : 1-11111111

Db 236 LPPNAGGGGGGGAGAPAI 254

RESULT 25

S20099

transforming protein jund - chicken

C:Species: Gallus gallus (chicken)

C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999

C:Accession: S20099

R:Hartl, M.; Hutchins, J.T.; Vogt, P.K.

Oncogene 6, 1623-1631, 1991

A:Title: The chicken jund gene and its product.

A:Reference number: S20099; MUID:92019832

A:Accession: S20099

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <HAR>

A:Cross-references: EMBL:X60063; NID:g62927; PIDN:CAA42665.1; PID:g62928

C:Superfamily: jun transforming protein; fos/jun DNA-binding domain homology

C:Keywords: DNA binding; nucleus; transcription regulation

F:237-277/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 31.6%; Score 60; DB 2; Length 323;

Best Local Similarity 72.2%; Pred. No. 16;

Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 11 AARAGGGGGGIEGPTL 28

1 : : 1-11111111

Db 151 AAAAGGGGGGGGGEL 168

RESULT 26

S58064

hdc protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 24-Sep-1998

C:Accession: S58064

R:Weaver, T.A.; White, R.A.

submitted to the EMBL Data Library, July 1995

A:Description: hdc, an imaginal specific gene required for adult morphogenesis in Drosophila

A:Reference number: S58064

A:Accession: S58064

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-649 <WEA>

A:Cross-references: EMBL:Z50097; NID:g902623; PID:g902624

C:Genetics:

A:Gene: FlyBase:hdc

A:Cross-references: FlyBase:FBgn0010113

Query Match 31.6%; Score 60; DB 2; Length 649;

Best Local Similarity 76.9%; Pred. No. 30;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPT 27

1 : : 1-11111111

Db 220 GGGGGGGGVNGNT 232

RESULT 27

T13386

hypothetical protein 115C2.3 - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13386; A39612

R:Salles, C.; Valenti, P.; Darlamitsou, A.; Henderson, N.; Campbell, L.; Glover, D.

submitted to the EMBL Data Library, May 1999

A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A:Reference number: Z17665

A:Accession: T13386

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1325 <CAT>

A:Cross-references: EMBL:AL031581; NID:el320978; PID:el320992; PIDN:CAA20886.1

R:Voelker, R.A.; Gibson, W.; Graves, J.P.; Sterling, J.F.; Eisenberg, M.T.

Mol. Cell. Biol. 11, 894-905, 1991

A:Title: The Drosophila suppressor of sable gene encodes a polypeptide with regions s

A:Reference number: A39612; MUID:91117256

A:Accession: A39612

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-719,723-1325 <VOE>

A:Cross-references: GB:M57889; NID:g158516; PID:g158517

C:Genetics:

A:Gene: FlyBase:su(s)

A:Cross-references: FlyBase:FBgn0003575; FlyBase:FBgn0020381

A:Map position: X
 A:Introns: 92/1; 170/3; 603/2; 645/1
 A:Note: EG:115C2.3

Query Match 31.6%; Score 60; DB 2; Length 1325;
 Best Local Similarity 68.8%; Pred. No. 56;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 15 GGGGGGGGIEGPTLRQ 30
 |||||
 Db 1162 GGGGGGGVLPNLSQ 1177

RESULT 28

E96495
 hypothetical protein F8D11.2 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 15-Jun-2001
 C:Accession: E96495
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 Hansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E96495
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-443 <STO>
 A:Cross-references: GB:AE005173; NID:gl0092398; PIDN:AAG12804.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F8D11.2
 A:Map position: 1
 C:Superfamily: barley pathogen resistance protein M10

Query Match 31.3%; Score 59.5; DB 2; Length 443;
 Best Local Similarity 92.3%; Pred. No. 24;
 Matches 12; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 14 AGGGGGGGIEGP 26
 |||||
 Db 2 AGGGGGGGG-EGP 13

RESULT 29

B39490
 subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form B - hum
 N:Alternate names: subtilisin homolog precursor, short splice form
 C:Species: Homo sapiens (man)
 C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 31-Mar-2000
 C:Accession: B39490
 R:Kiefer, M.C.; Tucker, J.E.; Joh, R.; Landsberg, K.E.; Saltman, D.; Barr, P.J.
 DNA Cell Biol. 10, 757-769, 1991
 A:Title: Identification of a second human subtilisin-like protease gene in the fes/fps
 A:Reference number: A39490; MUID:92075167
 A:Accession: B39490
 A:Molecule type: mRNA
 A:Residues: 1-487 <KIE>
 A:Note: the lack of a domain necessary for correct folding and activity of other serine
 C:Genetics:
 A:Gene: GDB:PACE4
 A:Cross-references: GDB:131390; OMIM:167405
 A:Map position: 15q26-15q26
 C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C:Keywords: alternative splicing; hydrolase; serine proteinase
 F:196-434/Domain: subtilisin homology <SBT>
 F:205,246,420/Active site: Asp, His, Ser #status predicted

Query Match 31.3%; Score 59.5; DB 2; Length 487;
 Best Local Similarity 60.0%; Pred. No. 27;
 Matches 15; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
 QY 11 AARAGGGGGGIEGPTLRQCLAAR 35
 |||||
 Db 24 AAGAGGAGGAGGAGGPGFRP-LAPR 47
 RESULT 30
 JC2191
 subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form C -
 N:Alternate names: kexin-like protease isoform
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 20-Apr-2000
 C:Accession: JC2191
 R:Tsuji, A.; Higashine, K.; Hine, C.; Mori, K.; Tamai, Y.; Nagamune, H.; Matsuda, Y.
 Biochem. Biophys. Res. Commun. 200, 943-950, 1994
 A:Title: Identification of novel cDNAs encoding human kexin-like protease, PACE4 isof
 A:Reference number: JC2191; MUID:94235049
 A:Accession: JC2191
 A:Molecule type: mRNA
 A:Residues: 1-652 <TSU>
 C:Comment: This protein consists of a signal peptide, a propeptide, a subtilisin-like
 C:Comment: This protein cleaves precursor proteins at dibasic amino acid residues.
 C:Genetics:
 A:Gene: GDB:PACE4
 A:Cross-references: GDB:131390; OMIM:167405
 A:Map position: 15q26-15q26
 C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
 C:Keywords: alternative splicing; hydrolase; serine proteinase
 F:196-434/Domain: subtilisin homology <SBT>
 F:205,246,420/Active site: Asp, His, Ser #status predicted
 Query Match 31.3%; Score 59.5; DB 1; Length 652;
 Best Local Similarity 60.0%; Pred. No. 34;
 Matches 15; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
 QY 11 AARAGGGGGGIEGPTLRQCLAAR 35
 |||||
 Db 24 AAGAGGAGGAGGAGGPGFRP-LAPR 47

Search completed: October 9, 2002, 09:05:06
 Job time : 9.09368 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2002, 08:51:41 ; Search time 4.29977 Seconds
(without alignments)
324.181 Million cell updates/sec

Title: US-09-422-838C-27

Perfect score: 190
Sequence: 1 IEQPTLRQCLARAGGGGGGEGTTLRQCLARA 36

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	38.4	619	1 LAC1_NEUCR	P06811 neurospora
2	73	38.4	619	1 LAC2_NEUCR	P10574 neurospora
3	65	34.2	105	1 INS_BOVIN	P01317 bos taurus
4	65	34.2	201	1 YR21_TRSVR	P25245 tomato ring
5	65	34.2	401	1 HB9_HUMAN	P50219 homo sapien
6	65	34.2	1263	1 SYV2_MOUSE	O92149 mus musculu
7	64.5	33.9	1733	1 VNUA_PVRA	P33485 pseudorabie
8	64	33.7	105	1 INS_SHEEP	P01318 ovis aries
9	64	33.7	584	1 CNAL_DROME	P12252 drosophila
10	63	33.2	394	1 FXD3_CHICK	P79772 gallus gall
11	63	33.2	1168	1 MYSC_ACACA	P10569 acanthamoeb
12	62	32.6	440	1 DCO_DROME	O76324 drosophila
13	61.5	32.4	4499	1 DTHA_CHLRE	O39610 chlamydomon
14	61	32.1	1178	1 PHYB_SORBI	P93527 sorghum bic
15	60	31.6	323	1 JUND_CHICK	P27921 gallus gall
16	60	31.6	348	1 SXL_CERCA	O61374 ceratitidis c
17	60	31.6	1322	1 SUS_DROME	P22293 drosophila
18	59.5	31.3	391	1 SOX1_MOUSE	P53783 mus musculu
19	59.5	31.3	969	1 PAC4_HUMAN	P29122 homo sapien
20	59	31.1	367	1 BET3_MESAU	O9029 mesocricetu
21	59	31.1	497	1 FXD2_HUMAN	O60548 homo sapien
22	59	31.1	517	1 Y967_TREPA	O83933 treponema p
23	59	31.1	593	1 K1CJ_HUMAN	P13645 homo sapien
24	59	31.1	753	1 Z1N_HUMAN	O9nr13 homo sapien
25	59	31.1	757	1 EGR_LUCU	O18531 lucilia cup
26	58.5	30.8	168	1 SSB_MTCLE	P46390 mycobacteri
27	58	30.5	445	1 HH3R_HUMAN	O9y5n1 homo sapien
28	58	30.5	476	1 EVX2_HUMAN	O3828 homo sapien
29	58	30.5	485	1 ONC2_HUMAN	O95948 homo sapien
30	58	30.5	495	1 BRN1_MOUSE	P31361 mus musculu
31	58	30.5	497	1 BRN1_RAT	O63262 rattus norv
32	58	30.5	500	1 BRN1_HUMAN	P20264 homo sapien
33	58	30.5	569	1 K1CJ_MOUSE	P02535 mus musculu

ALIGNMENTS

RESULT 1

ID	LAC1_NEUCR	STANDARD;	PRT;	619 AA.
AC	P06811;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Laccase precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)			
DE	(Urishiol oxidase) (Laccase allele OR).			
GN	LACC.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Sordariales; Sordariaceae; Neurospora.			
NCBI_TaxID=5141;				
ON	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=88087214; PubMed=2961749;			
RA	Germann U.A., Mueller G., Hunziker P.E., Lerch K.;			
RT	"Characterization of two allelic forms of Neurospora crassa laccase.			
RT	Amino- and carboxyl-terminal processing of a precursor.";			
RL	J. Biol. Chem. 263:885-896(1988).			
RL	[2]			
RP	SEQUENCE OF 379-619 FROM N.A.			
RX	MEDLINE=87067412; PubMed=2947240;			
RA	Germann U.A., Lerch K.;			
RT	"Isolation and partial nucleotide sequence of the laccase gene from			
RT	Neurospora crassa: amino acid sequence homology of the protein to			
RT	human ceruloplasmin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:8854-8858(1986).			
CC	- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED			
CC	PRODUCTS (PROBABLE).			
CC	- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2			
CC	H(2)O.			
CC	- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU			
CC	CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE			
CC	3 OR COUPLED BINUCLEAR (BY SIMILARITY).			
CC	- SUBCELLULAR LOCATION: Secreted (Potential).			
CC	- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.			
CC	- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M14554; AAA33590.1; -			
DR	EMBL; M18333; AAA33591.1; -			
DR	PIR; A28523; KNCLO.			
DR	PIR; A29762; A29762.			
DR	InterPro; IPR001117; Cu-oxidase.			
DR	InterPro; IPR002355; MultiCu_Oxidase2.			
DR	Pfam; PF00394; Cu-Oxidase; 3.			
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.			

P54865 cellulomona
O54839 mus musculu
O55165 rattus norv
P25764 oryza sativ
Q15911 homo sapien
P49842 homo sapien
O84537 rattus norv
O84556 mus musculu
P17656 caenorhabdi
P23091 avian muscu
P81705 drosophila
Q98937 gallus gall

DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 KW Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 49 LACCASE.
 FT CHAIN 50 606
 FT PROPEP 607 619
 FT DOMAIN 84 207 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 216 373 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 431 566 PLASTOCYANIN-LIKE 3.
 FT METAL 144 144 COPPER (TYPE 2) (PROBABLE).
 FT METAL 146 146 COPPER (TYPE 3) (PROBABLE).
 FT METAL 189 189 COPPER (TYPE 3) (PROBABLE).
 FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
 FT METAL 477 477 COPPER (TYPE 1) (PROBABLE).
 FT METAL 480 480 COPPER (TYPE 2) (PROBABLE).
 FT METAL 482 482 COPPER (TYPE 3) (PROBABLE).
 FT METAL 548 548 COPPER (TYPE 3) (PROBABLE).
 FT METAL 549 549 COPPER (TYPE 1) (PROBABLE).
 FT METAL 550 550 COPPER (TYPE 1) (PROBABLE).
 FT METAL 554 554 COPPER (TYPE 1) (PROBABLE).
 FT METAL 559 559 COPPER (TYPE 1) (PROBABLE).
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 619 AA; FDED6D78B65048E3 CRC64;
 Query Match 38.4%; Score 73; DB 1; Length 619;
 Best Local Similarity 60.0%; Pred. No. 0.64;
 Matches 15; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 7 RCLAAAGGGGGGGGIEGPTLRQC 31
 || | ||||| || |||
 Db 39 RQDSQAERYGGGGGCGNSPTNRQC 63
 RESULT 2
 LAC2_NEUCR LAC2_NEUCR STANDARD; PRT; 619 AA.
 AC P10574;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Laccase precursor (EC 1.10.3.2) (Benzenediol: oxygen oxidoreductase)
 DE (Urishiol oxidase) (Laccase allele TS).
 GN LACC.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88087214; PubMed=2961749;
 RA Germann U.A., Mueller G., Hunziker P.E., Lerch K.;
 RT "Characterization of two allelic forms of Neurospora crassa laccase.
 RT Amino- and carboxyl-terminal processing of a precursor.";
 RL J. Biol. Chem. 263:885-896(1988).
 CC -1- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
 CC PRODUCTS (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: 4 benzenediol + O(2) -> 4 benzosemiquinone + 2
 CC H(2)O.
 CC -1- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
 CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
 CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
 CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL; M18334; AAA33592.1; -.
 DR PIR; B28523; KSNCLT.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR002355; MultiCu_oxidase2.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
 DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
 KW Oxidoreductase; Signal; Copper; Metal-binding; Lignin degradation;
 KW Glycoprotein; Repeat.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 49
 FT CHAIN 50 606 LACCASE.
 FT PROPEP 607 619
 FT DOMAIN 84 207 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 216 373 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 431 566 PLASTOCYANIN-LIKE 3.
 FT METAL 144 144 COPPER (TYPE 2) (PROBABLE).
 FT METAL 146 146 COPPER (TYPE 3) (PROBABLE).
 FT METAL 189 189 COPPER (TYPE 3) (PROBABLE).
 FT METAL 191 191 COPPER (TYPE 3) (PROBABLE).
 FT METAL 477 477 COPPER (TYPE 1) (PROBABLE).
 FT METAL 480 480 COPPER (TYPE 2) (PROBABLE).
 FT METAL 482 482 COPPER (TYPE 3) (PROBABLE).
 FT METAL 548 548 COPPER (TYPE 3) (PROBABLE).
 FT METAL 549 549 COPPER (TYPE 1) (PROBABLE).
 FT METAL 550 550 COPPER (TYPE 3) (PROBABLE).
 FT METAL 554 554 COPPER (TYPE 1) (PROBABLE).
 FT METAL 559 559 COPPER (TYPE 1) (PROBABLE).
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 619 AA; 68120 MW; 0B86CCDE18841145 CRC64;
 Query Match 38.4%; Score 73; DB 1; Length 619;
 Best Local Similarity 60.0%; Pred. No. 0.64;
 Matches 15; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 7 RCLAAAGGGGGGGGIEGPTLRQC 31
 || | ||||| || |||
 Db 39 RQDSQAERYGGGGGCGNSPTNRQC 63
 RESULT 3
 INS_BOVIN INS_BOVIN STANDARD; PRT; 105 AA.
 AC P01317;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Insulin precursor.
 DE INS.
 GN INS.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88288209; PubMed=2456452;
 RA D'Agostino J., Younes M.A., White J.W., Besch P.K., Field J.B.,
 RA Frazier M.L.;
 RT "Cloning and nucleotide sequence analysis of complementary
 RT deoxyribonucleic acid for bovine preproinsulin.";
 RL Mol. Endocrinol. 1:327-331(1987).

FT DOMAIN 144 148 POLY-GLY.
 SQ SEQUENCE 201 AA; 20194 MW; 9038506E18D7B450 CRC64;
 Query Match 34.2%; Score 65; DB 1; Length 201;
 Best Local Similarity 61.5%; Pred. No. 1.7; Mismatches 5; Indels 4; Gaps 1;
 Matches 16; Conservative 1;

QY 13 RAGGGGGGGGIE---GPTLRQCLAA 34
 |||||
 DB 13 RAGGGGGGGGKEVFRAGRLLKVLKA 38
 |||||

RESULT 5
 HB9_HUMAN STANDARD; PRT; 401 AA.
 AC P50219;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Homeobox protein HB9.
 GN HLBX9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=94327547; PubMed=7914194;
 RA Harrison K.A., Druey K.M., Deguchi Y., Tuscano J.M., Kehrl J.H.;
 RT "A novel human homeobox gene distantly related to proboscipedia is
 expressed in lymphoid and pancreatic tissues.";
 RL J. Biol. Chem. 269:19968-19975(1994).
 CC - FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.
 CC - SIMILARITY: TO DROSOPHILA HOMEOBOX PROTEIN PROBOSCIPEDIA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U07664; AAB60647.1;
 DR EMBL; U07663; AAB60647.1; JOINED.
 DR HSP; P14653; I872.
 DR TRANSFAC; T03420; -.
 DR MIM; 142994; -.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox.1.
 DR PRINTS; PR00024; HOMEOBOX.
 DR SMART; SM00389; HOX.1.
 DR PROSITE; PS00027; HOMEOBOX_1; 1.
 DR PROSITE; PS50071; HOMEOBOX_2; 1.
 KW Homeobox; DNA-binding; Nuclear protein; Transcription regulation.
 FT DOMAIN 39 48
 FT POLY-GLY.
 FT DOMAIN 97 111
 FT POLY-GLY.
 FT DOMAIN 120 135
 FT POLY-ALA.
 FT DOMAIN 169 177
 FT POLY-ALA.
 FT DNA_BIND 242 301
 FT HOMEOBOX.
 FT DOMAIN 316 325
 FT POLY-GLY.
 SQ SEQUENCE 401 AA; 40932 MW; 0006AEAD71D594FE CRC64;

Query Match 34.2%; Score 65; DB 1; Length 401;
 Best Local Similarity 54.2%; Pred. No. 3; Mismatches 10; Indels 0; Gaps 0;
 Matches 13; Conservative 1;

QY 11 AARAGGGGGGGGTEGPTLRQCLAA 34
 |||||
 DB 37 ASGTGGGGGGGASGSGTSCSPA 60
 |||||

RESULT 6
 SYV2_MOUSE STANDARD; PRT; 1263 AA.
 AC Q921Q9; Q9QUN2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Valyl-tRNA synthetase 2 (EC 6.1.1.9) (Valine--tRNA ligase 2) (VALRS
 2).
 DE VARS2 OR G7A OR BAT6.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RA Rowen L., Madan A., Qin S., Shaffer T., Ratcliffe A., Abbasi N.,
 RA Dickhoff R., James R., Loretz C., Lasky S., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 region.";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/C, AND C57BL/RIJ; TISSUE=Brain;
 RX MEDLINE=99216447; PubMed=10199925;
 RA Snek M., van Vugt H.;
 RT "The sequence and organization of the mouse valyl-tRNA synthetase gene
 G7a/bat6 located in the MHC class III region.";
 RL Immunogenetics 49:468-470(1999).
 CC - CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
 CC + L-valyl-tRNA(Val).
 CC - ENZYME REGULATION: CAN BE REGULATED BY PROTEIN KINASE C-DEPENDENT
 CC PHOSPHORYLATION (BY SIMILARITY).
 CC - SUBUNIT: FORMS HIGH-MOLECULAR-MASS AGGREGATES WITH ELONGATION
 CC FACTOR 1 (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC - SIMILARITY: THE N-TERMINAL DOMAIN IS SIMILAR TO ELONGATION
 CC FACTOR 1-GAMMA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF109905; AAC84151.1; -.
 DR EMBL; AF109906; AAC84172.1; -.
 DR EMBL; AF087680; AAD26532.1; -.
 DR EMBL; AF087141; AAD26531.1; -.
 DR HSP; P96142; ICAX.
 DR MGD; MGI:90875; Vars2.
 DR InterPro; IPR004046; GST-C.
 DR InterPro; IPR004045; GST-N.
 DR InterPro; IPR002300; tRNA-synt_1a.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002303; tRNA-synt_val.
 DR Pfam; PF00043; GST-C; 1.
 DR Pfam; PF02798; GST-N; 1.
 DR Pfam; PF00133; tRNA-synt_1; 1.
 DR PRINTS; PS00986; TRNASYNTHAL.
 DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 Phosphorylation.
 FT DOMAIN 1 7200 EF-1-GAMMA LIKE.
 FT SITE 343 353 "HIGH" REGION.
 FT SITE 861 863 "KMSK" REGION.
 FT BINDING 864 864 ATP (BY SIMILARITY).
 FT CONFLICT 959 959 A -> R (IN REF. 2).

```

FT CONFLICT 1219 1219 E -> K (IN REF. 2).
SQ SEQUENCE 1263 AA; 140214 MW; B510E73284FCE26D CRC64;

Query Match
Best Local Similarity 34.2%; Score 65; DB 1; Length 1263;
Matches 15; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 4 PTLROCLAAARAGGGGGGIEGPTLROCL 32
   :||| :||| :||| :||| :||| :|||
Db 14 PSLRALIAARYGEAGDGPWGPHPRICL 42

RESULT 7
VNUA_PRVKA
ID VNUA_PRVKA STANDARD; PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DE 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable nuclear antigen.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vleck C., Kozmik Z., Paces V., Schirm S., Schwyzler M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M34651; AAA47471.1; -
DR PIR; B45344; B45344.
DR DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475B85E2 CRC64;

Query Match
Best Local Similarity 33.9%; Score 64.5; DB 1; Length 1733;
Matches 18; Conservative 2; Mismatches 13; Indels 9; Gaps 2;

Qy 3 GPTLROCL-AARAGGGGGG-----GGIEGPTLROCLAAAR 35
   ||| :||| :||| :||| :||| :|||
Db 1645 GPSPRGCRGAGRAGGGGGCGGGRAGPAGAGGGLRCRCCR 1686

RESULT 8
INS_SHEEP
ID INS_SHEEP STANDARD; PRT; 105 AA.
AC P01318;
DT 21-JUL-1986 (Rel. 01, Created)
DE 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin precursor.
GN INS.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoides;

```

```

OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94280618; PubMed=8011164;
RA Ohlsen S.M., Lugenbeel K.A., Wong E.A.;
RT "Characterization of the linked ovine insulin and insulin-like growth
RT factor-II genes.";
RL DNA Cell Biol. 13:377-388(1994).
RN [2]
RP SEQUENCE OF 25-54 AND 85-105.
RA Brown H., Sanger F., Kitai R.;
RT "The structure of pig and sheep insulins.";
RL Biochem. J. 60:556-565(1955).
RN [3]
RP SEQUENCE OF 57-82.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
RT dog proinsulin C-peptides by a semi-micro Edman degradation
RT procedure";
RL J. Biol. Chem. 247:4866-4871(1972).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00659; AAB60625.1; -
DR PIR; S16430; INSH.
DR HSSP; P01315; 9INS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 82 C PEPTIDE.
FT CHAIN 85 105 INSULIN A CHAIN.
FT DISULFID 31 91 INTERCHAIN.
FT DISULFID 43 104 INTERCHAIN.
FT DISULFID 90 95
SQ SEQUENCE 105 AA; 11235 MW; 8B27C7FB9922BC7A CRC64;

Query Match
Best Local Similarity 33.7%; Score 64; DB 1; Length 105;
Matches 16; Conservative 3; Mismatches 11; Indels 8; Gaps 2;

Qy 1 IEGPLRQCLAAARAGGGGGGIEGP-----TLROCLA 33
   :||| :||| :||| :||| :||| :|||
Db 58 VEGP---QVGALEAGPGAGGLEGPQKRGIVEQCCA 92

RESULT 9
CNAL_DROME
ID CNAL_DROME STANDARD; PRT; 584 AA.
AC P12252;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```


FT NP_BIND 15 23 ATP (BY SIMILARITY).
 FT BINDING 38 38 ATP (BY SIMILARITY).
 FT ACT_SITE 128 128 BY SIMILARITY.
 FT DOMAIN 319 332 POLY-ALA.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 351 POLY-GLY.
 FT DOMAIN 414 426 POLY-GLY.
 FT DOMAIN 430 437 POLY-GLY.
 FT MUTAGEN 47 47 P-S: IN DBTS; SHORTENS THE BEHAVIORAL PERIOD.
 FT MUTAGEN 80 80 M-SI: IN DBTL; LENGTHENS THE BEHAVIORAL PERIOD.
 FT SEQUENCE 440 AA; 48073 MW; B875891D5747391D CRC64;
 SQ
 Query Match 32.6%; Score 62; DB 1; Length 440;
 Best Local Similarity 55.0%; Pred. No. 6.6;
 Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 Qy 4 PTLRQCLAARAGGGGGGGI 23
 Db 403 PRRPSIRMGGGGGGGV 422
 RESULT 13
 DYHA_CHLRE STANDARD; PRT; 4499 AA.
 AC Q39610;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dynein alpha chain, flagellar outer arm (DHC alpha).
 GN ODA11 OR ODA-11.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A., AND REVISIONS.
 RC STRAIN=21GR;
 RA MEDLINE=97329535; PubMed=9186009;
 PA Mitchell D.R., Brown K.S.:
 RT "Sequence analysis of the Chlamydomonas reinhardtii flagellar alpha dynein gene";
 RL Cell Motil. Cytoskeleton 37:120-126(1997).
 RN [2]
 RP SEQUENCE OF 1142-4499 FROM N.A.
 RC STRAIN=21GR;
 RX MEDLINE=94274778; PubMed=8006077;
 RA Mitchell D.R., Brown K.S.:
 RT "Sequence analysis of the Chlamydomonas alpha and beta dynein heavy chain genes";
 RL J. Cell Sci. 107:635-644(1994).
 CC -!- FUNCTION: FORCE GENERATING PROTEIN OF EUKARYOTIC CILIA AND FLAGELLA. PRODUCES FORCE TOWARDS THE MINUS ENDS OF MICROTUBULES. DYNEIN HAS ATPASE ACTIVITY.
 CC -!- SUBUNIT: CONSISTS OF AT LEAST 3 HEAVY CHAINS (ALPHA, BETA AND GAMMA), 2 INTERMEDIATE CHAINS AND 8 LIGHT CHAINS.
 CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L26049; AAA57316.2;
 CC InterPro; IPR003593; AAA.
 CC InterPro; IPR001298; Filamin.
 CC InterPro; IPR002909; IPT_TIG.
 CC InterPro; IPR001798; Kelch.
 CC InterPro; IPR001736; PLD.

DR Pfam: PF00630; Filamin; 1.
 DR Pfam: PF01344; Kelch; 3.
 DR SMART: SM00382; AAA; 3.
 DR SMART: SM00429; IPT; 1.
 DR PROSITE: PS50194; FILAMIN_REPEAT; 1.
 KW Motor protein; Microtubules; Dynein; ATP-binding; Flagella;
 KW Coiled coil.
 FT REPEAT 425 534 FILAMIN.
 FT DOMAIN 1261 1334 COILED COIL (POTENTIAL).
 FT DOMAIN 1382 1450 COILED COIL (POTENTIAL).
 FT DOMAIN 1836 1864 MICROTUBULE-BINDING (POTENTIAL).
 FT DOMAIN 2655 2688 COILED COIL (POTENTIAL).
 FT DOMAIN 3003 3023 COILED COIL (POTENTIAL).
 FT DOMAIN 3170 3262 COILED COIL (POTENTIAL).
 FT DOMAIN 3486 3515 COILED COIL (POTENTIAL).
 FT NP_BIND 1716 1723 ATP (POTENTIAL).
 FT NP_BIND 2019 2026 ATP (POTENTIAL).
 FT NP_BIND 2369 2376 ATP (POTENTIAL).
 FT NP_BIND 2717 2754 ATP (POTENTIAL).
 SQ SEQUENCE 4499 AA; 503606 MW; 319AC7FD30F1591A CRC64;
 Query Match 32.4%; Score 61.5; DB 1; Length 4499;
 Best Local Similarity 48.5%; Pred. No. 54;
 Matches 16; Conservative 3; Mismatches 11; Indels 3; Gaps 1;
 Qy 3 GPTLRQCLAARAGGGGGGIEG---PTLRQCL 32
 Db 4194 GETLFKTVVEVAGGGGGGGGGGGENAVRQAL 4226
 RESULT 14
 PHYB_SORBI STANDARD; PRT; 1178 AA.
 AC P93527;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Phytochrome B.
 GN PHYB OR MA3.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4538;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. 58M;
 RX MEDLINE=20188796; PubMed=10723737;
 RA Alba R., Kelmanson P.M., Cordonnier-Pratt M.-M., Pratt L.H.:
 RT "The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms";
 RL Mol. Biol. Evol. 17:362-373(2000).
 RN [2]
 RP SEQUENCE OF 208-1178 FROM N.A.
 RC STRAIN=CV. 58M;
 RX MEDLINE=97198556; PubMed=9046599;
 RA Childs K.L., Miller F.R., Cordonnier-Pratt M.-M., Pratt L.H., Morgan P.W., Mullet J.E.;
 RT "The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a phytochrome B";
 RL Plant Physiol. 113:611-619(1997).
 CC -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RUBULOSE-BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN, PROTOCHLOROPHYLLIDE REDUCTASE, RRNA, ETC. IT ALSO CONTROLS THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION (BY SIMILARITY).

CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF182394; BAB41398.2; -
DR InterPro: IPR003018; GAF; -
DR InterPro: IPR003594; HATPase_c.
DR InterPro: IPR004359; HIS_KIN_sig.
DR InterPro: IPR003661; HIS_kinA.
DR InterPro: IPR000014; PAS.
DR InterPro: IPR001294; Phytochrome.
DR Pfam: PF01590; GAF; 1.
DR Pfam: PF02318; HATPase_c; 1.
DR Pfam: PF00989; PAS; 2.
DR Pfam: PF00360; phytochrome; 1.
DR Pfam: PF00512; signal; 1.
DR PRINTS: PR01033; PHYTOCHROME.
DR SMART: SM00065; GAF; 1.
DR SMART: SM00387; HATPase_c; 1.
DR SMART: SM00388; HSKA; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS50109; HIS_KIN; 1.
DR PROSITE: PS50112; PAS; 2.
DR PROSITE: PS00245; PHYTOCHROME_1; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat; Multigene family.
FT DOMAIN 668 739 PAS 1.
FT DOMAIN 802 873 PAS 2.
FT DOMAIN 950 1170 HISTIDINE KINASE.
FT DOMAIN 23 31 POLY-HIS.
FT DOMAIN 43 54 POLY-GLY.
FT BINDING 372 372 CHROMOPHORE (BY SIMILARITY).
SQ SEQUENCE 1178 AA; 129136 MW; C406DF221197B93F CRC64;

Query Match 32.1%; Score 61; DB 1; Length 1178;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 12 ARAGGGGGGGGIEGPT 27
Db 40 SRAGGGGGGGGGGGT 55
:|||||
RESULT 15
JUND_CHICK STANDARD; PRT; 323 AA.
AC P27921;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Transcription factor jund-D.
GN JUND.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92019832; PubMed=1923529;
RA Hartl M., Hutchins J.T., Vogt P.K.;
RT "The chicken jund gene and its product.";

RL Oncogene 6:1623-1631(1991).
CC -!- SUBUNIT: BINDS DNA AS A DIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. JUN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60063; CAA42665.1; -
DR PIR: S20099; S20099.
DR HSSP: P05412; LFOS.
DR TRANSFAC: T02196; -
DR InterPro: IPR002112; Leuzip_Jun.
DR InterPro: IPR001871; BZIP.
DR Pfam: PF00170; bzip; 1.
DR PRINTS: PR00043; LEUZIPPJUN.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS00036; BZIP_BASIC; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 59 67 POLY-ALA.
FT DOMAIN 155 166 POLY-GLY.
FT DNA_BIND 242 266 BASIC MOTIF.
FT DOMAIN 270 298 LEUCINE-ZIPPER.
SQ SEQUENCE 323 AA; 33205 MW; A7F6D21A97DBB676 CRC64;

Query Match 31.6%; Score 60; DB 1; Length 323;
Best Local Similarity 72.2%; Pred. No. 8.2;
Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 11 AARAGGGGGGGGIEGPTL 28
Db 151 AAAAGGGGGGGGGGEL 168
:|||||
RESULT 16
SXL_CERCA STANDARD; PRT; 348 AA.
ID SXL_CERCA
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sex-lethal protein homolog (CCSXL).
GN SXL.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Ceratitis.
OX NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GENAKIO;
RA MEDLINE=98171464; PubMed=9502730;
RA Saccone G., Peluso I., Artiaco D., Giordano E., Bopp D., Polito L.C.;
RT "The Ceratitis capitata homologue of the Drosophila sex-determining
RT gene sex-lethal is structurally conserved, but not sex-specifically
RT regulated.";
RL Development 125:1495-1500(1998).
CC -!- FUNCTION: UNKNOWN; APPARENTLY NOT INVOLVED IN SOMATIC SEX
CC DETERMINATION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS; ADULT-SPECIFIC ISOFORMS
CC A1, A2, A3, A4, AND EMBRYO-SPECIFIC ISOFORMS E1, E2 AND E3 (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN EMBRYOS OF BOTH SEXES. ALSO
CC EXPRESSED IN THE PROGENITOR CELLS OF THE GERMLINE.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF026145; AAC38968.1; -
 DR HSSP; P19339; ISXL
 DR InterPro; IPR000504; RRM.
 DR Pfam; PF00076; rrm; 2.
 DR PRINTS; PR00961; HUDSXLRNA.
 DR SMART; SM00360; RRM; 2.
 DR PROSITE; PS0102; RRM; 2.
 DR PROSITE; PS0030; RRM_RNP.1; 1.
 KW RNA-binding; Repeat; Nuclear protein; Alternative splicing.
 FT DOMAIN 1 27 GLY/ASN-RICH DOMAIN.
 FT DOMAIN 110 188 RNA-BINDING (RRM) 1.
 FT DOMAIN 196 276 RNA-BINDING (RRM) 2.
 FT DOMAIN 68 75 POLY-GLY.
 FT DOMAIN 95 99 POLY-GLY.
 FT DOMAIN 293 311 POLY-GLY.
 FT DOMAIN 312 316 POLY-PRO.
 FT VARSPLIC 37 44 MISSING (IN ISOFORM AL).
 SQ SEQUENCE 348 AA; 37188 MW; CAB3AD5C2C8874A CRC64;

Query Match 31.6%; Score 60; DB 1; Length 348;
 Best Local Similarity 83.3%; Pred. No. 8.7;
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGP 26

|||||||: ||

DB 301 GGGGGGGGGMGP 312

RESULT 17

SUS_DROME STANDARD; PRT; 1322 AA.
 ID SUS_DROME
 AC P22293;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Suppressor of sable protein.
 GN SU(S).
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=91117256; PubMed=1703632;
 RA Voelker R.A., Gibson W., Graves J.P., Sterling J.F., Eisenberg M.T.;
 RT "The Drosophila suppressor of sable gene encodes a polypeptide with
 RT regions similar to those of RNA-binding proteins.";
 RL Mol. Cell. Biol. 11:894-905(1991).
 RN [2]
 RP SEQUENCE OF 1-9 FROM N.A.
 RX MEDLINE=91169252; PubMed=1963868;
 RA Voelker R.A., Graves J.P., Gibson W., Eisenberg M.T.;
 RT "Mobile element insertions causing mutations in the Drosophila
 RT suppressor of sable locus occur in DNase I hypersensitive subregions
 RT of 5'-transcribed nontranslated sequences.";
 RL Genetics 126:1071-1082(1990).
 CC -1- FUNCTION: AFFECTS THE TRANSCRIPT LEVELS OF THOSE ALLELES THAT IT
 CC SUPPRESSES. MAY BE INVOLVED IN RNA METABOLISM.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: AT ALL STAGES.
 CC -1- SIMILARITY: HAS REGIONS SIMILAR TO THOSE OF RNA-BINDING PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M57889; AAA28920.1; -
 DR EMBL; X59364; CAA42010.1; -
 DR PIR; A39612; A39612.
 DR FlyBase; FBgn0003575; su(s).
 DR InterPro; IPR000571; Zf-CCCH.
 DR Pfam; PF00642; Zf-CCCH; 2.
 KW RNA-binding; Nuclear protein.
 FT DOMAIN 138 327 HIGHLY CHARGED DOMAIN.
 FT DOMAIN 446 474 GLN-RICH (OPA-REPEAT).
 FT DOMAIN 1087 1162 RNA-BINDING (BY SIMILARITY).
 SQ SEQUENCE 1322 AA; 143555 MW; D5F534EB5702EA08 CRC64;

Query Match 31.6%; Score 60; DB 1; Length 1322;
 Best Local Similarity 68.8%; Pred. No. 27;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 15 GGGGGGGGIEGPTLRQ 30

|||||||: ||

DB 1159 GGGGGGGGVLPNLSQ 1174

RESULT 18

SOX1_MOUSE STANDARD; PRT; 391 AA.
 ID SOX1_MOUSE
 AC P53783;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SOX-1 protein.
 GN SOX1 OR SOX-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=96189340; PubMed=8625802;
 RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,
 RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N.,
 RA Lovell-Badge R.;
 RT "A comparison of the properties of Sox-3 with Sry and two related
 RT genes, Sox-1 and Sox-2.";
 RL Development 122:509-520(1996).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: MAINLY IN THE DEVELOPING CENTRAL NERVOUS
 CC SYSTEM. EXPRESSED IN DEVELOPING UROGENITAL RIDGE.
 CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X94126; CAA63846.1; -
 DR HSSP; Q05056; 1HRX
 DR MGD; MGI:98357; Sox1.
 DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 1.
 DR SMART; SM00398; HMG; 1.
 KW DNA-binding; Nuclear protein.
 FT DOMAIN 30 43
 FT DOMAIN 51 119 HMG_BOX.
 FT DOMAIN 145 150 POLY-GLY.

FT DOMAIN 197 204 POLY-ALA.
 FT DOMAIN 280 288 POLY-ALA.
 FT DOMAIN 296 306 POLY-ALA.
 FT DOMAIN 357 364 POLY-ALA.
 SQ SEQUENCE 391 AA; 392337 MW; 9F81ED667F947C05 CRC64;
 Query Match 31.3%; Score 59.5; DB 1; Length 391;
 Best Local Similarity 54.5%; Pred. No. 11;
 Matches 12; Conservative 1; Mismatches 4; Indels 5; Gaps 1;
 QY 1 IEPTLRQCLAAAGGGGGGGG 22
 Db 22 LSGPA-----GARGGGGGGGG 38
 RESULT 19
 PACE4_HUMAN
 ID PAC4_HUMAN STANDARD; PRT; 969 AA.
 AC P29122; Q15099; Q15100; Q9UEJ1; Q9UEJ2; Q9UEJ7; Q9UEJ8; Q9UEJ9;
 AC Q9UEG7; Q9Y4G9; Q9Y4H0; Q9Y4H1;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
 DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like proprotein
 DE convertase 4) (SPC4).
 GN PACE4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).
 RC TISSUE-Hepatoma, and Kidney;
 RX MEDLINE=92075167; PubMed=1741956;
 RA Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,
 RA Barr P.J.;
 RA "Identification of a second human subtilisin-like protease gene in
 RT the fes/fps region of chromosome 15";
 RL DNA Cell Biol. 10:757-769(1991).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).
 RC TISSUE-Placenta;
 RX MEDLINE=94235049; PubMed=8179631;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms";
 RL Biochem. Biophys. Res. Commun. 200:943-950(1994).
 RN [3]
 RP ERRATUM.
 RX MEDLINE=95071480; PubMed=7980617;
 RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,
 RA Matsuda Y.;
 RT "Identification of novel cDNAs encoding human kexin-like protease,
 RT PACE4 isoforms";
 RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM PACE4A-II).
 RC TISSUE-Placenta;
 RA Mori K., Imamaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,
 RA Matsuda Y.;
 RT "Identification of a novel PACE4 isoform, PACE4E";
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).
 RC TISSUE-Cerebellum;
 RX MEDLINE=97335942; PubMed=9192737;
 RA Mori K., Kii S., Tsuji A., Nagahama M., Imamaki A., Hayashi K.,
 RA Akamatsu T., Nagamune H., Matsuda Y.;
 RT "A novel human PACE4 isoform, PACE4E is an active processing protease
 RT containing a hydrophobic cluster at the carboxy terminus";
 RL J. Biochem. 121:941-948(1997).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORMS PACE4A-I; A-II; CS; D; E-I; E-II).
 RX MEDLINE=98021085; PubMed=9378725;
 RA Tsuji A., Hine C., Tamai Y., Yonemoto K., Mori K., Yoshida S.,
 RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
 RT "Genomic organization and alternative splicing of human PACE4 (SPC4),
 RT kexin-like processing endoprotease";
 RL J. Biochem. 122:438-452(1997).
 RN [7]
 RP ALTERNATIVE SPLICING (ISOFORM PACE4CS).
 RX MEDLINE=97064242; PubMed=8906861;
 RA Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
 RT "Functional analysis of human PACE4-A and PACE4-C isoforms:
 RT identification of a new PACE4-CS isoform";
 RL FEBS Lett. 396:31-36(1996).
 RN [8]
 RP CHARACTERIZATION.
 RX MEDLINE=9923559; PubMed=10215603;
 RA Sucic J.F., Moehring J.M., Innocencio N.M., Luchini J.W.,
 RA Moehring T.J.;
 RT "Endoprotease PACE4 is Ca2+-dependent and temperature-sensitive and
 RT can partly rescue the phenotype of a furin-deficient cell strain";
 RL Biochem. J. 339:639-647(1999).
 RN [9]
 RP PROCESSING.
 RX MEDLINE=98408849; PubMed=9738469;
 RA Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
 RA Tsuji A., Matsuda Y.;
 RT "Biosynthetic processing and quaternary interactions of proprotein
 RT convertase SPC4 (PACE4)";
 RL FEBS Lett. 434:155-159(1998).
 CC [1] FUNCTION: LIKELY TO REPRESENT AN ENDOPEPTIDASE ACTIVITY WITHIN THE
 CC CONSTITUTIVE SECRETORY PATHWAY, WITH UNIQUE RESTRICTED
 CC DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
 CC AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC [2] CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROTEIN BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS,
 CC WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC [3] COFACTOR: PACE4A IS PROBABLY CALCIUM-DEPENDENT.
 CC [4] SUBUNIT: THE PACE4A-I PRECURSOR PROTEIN SEEMS TO EXIST IN THE
 CC RETICULUM ENDOPLASMIC AS BOTH A MONOMER AND A DIMER-SIZED COMPLEX
 CC WHEREAS MATURE PACE4A-I EXISTS ONLY AS A MONOMER, SUGGESTING THAT
 CC PROPEPTIDE CLEAVAGE AFFECTS ITS TERTIARY OR QUATERNARY STRUCTURE.
 CC [5] SUBCELLULAR LOCATION: PACE4A-I AND PACE4A-II ARE SECRETED. PACE4C
 CC AND PACE4CS ARE NOT SECRETED AND REMAIN PROBABLY IN ZYMOGEN FORM
 CC IN ENDOPLASMIC RETICULUM. PACE4E-I AND PACE4E-II ARE RETAINED
 CC INTRACELLULARLY PROBABLY THROUGH A HYDROPHOBIC CLUSTER IN THEIR C-
 CC TERMINUS. PACE4B MIGHT BE SECRETED.
 CC [6] ALTERNATIVE PRODUCTS: 8 ISOFORMS: PACE4A-I/PACE4 (SHOWN HERE),
 CC PACE4A-II, PACE4B/PACE4.1, PACE4C, PACE4CS, PACE4D, PACE4E-I AND
 CC PACE4E-II; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS PACE4B,
 CC C, CS AND D MIGHT BE ENZYMATICALLY INACTIVE.
 CC [7] TISSUE SPECIFICITY: EACH PACE4 ISOFORM EXHIBITS A UNIQUE
 CC RESTRICTED DISTRIBUTION. PACE4A-I IS EXPRESSED IN HEART, BRAIN,
 CC PLACENTA, LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS, BUT AT
 CC COMPATIVELY HIGHER LEVELS IN THE LIVER. PACE4A-II IS AT LEAST
 CC EXPRESSED IN PLACENTA. PACE4B WAS ONLY FOUND IN THE EMBRYONIC
 CC KIDNEY CELL LINE FROM WHICH IT WAS ISOLATED. PACE4C AND PACE4D ARE
 CC EXPRESSED IN PLACENTA. PACE4E-I IS EXPRESSED IN CEREBELLUM,
 CC PLACENTA AND PITUITARY. PACE4E-II IS AT LEAST PRESENT IN
 CC CEREBELLUM.
 CC [8] DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM. ISOFORM PACE4D LACKS THE PROPEPTIDE DOMAIN.
 CC [9] SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC [10] SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

Seq. Local Similarity	Seq. no.	Seq. no.
Matches 15; Conservative 0; Mismatches 9; Indels 1; Gaps 1;		

QY	11	AARAGGGGGG	IEGPTLRQCL	AAR	35
Db	24	AAGAGGAGG	AGGAGGCGFR	P-LAPR	47

RESULT 20		
BET3_MESAU		
ID	BET3_MESAU	STANDARD; PRT; 367 AA.
AC	09029;	
DT	30-MAY-2000	(Rel. 39, Created)
DT	30-MAY-2000	(Rel. 39, Last sequence update)
DT	16-OCT-2001	(Rel. 40, Last annotation update)

OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OC NCBI TaxID=10036.

SEQUENCE FROM N.A.
MEDLINE=961140430; PubMed=8552091;
RAY Peyton M., Steilrecht C.M.M., Naya F.J., Huang H.-p., Samora P.J.,
Tsai M.-J.;
RT "BETA3", a novel helix-loop-helix protein, can act as a negative
regulator of BETA2 and MyoD-responsive genes.";
Mol. Cell. Biol. 16:626-633(1996).

CC		-I- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3 (E47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF NEUROD1 AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.
CC		-I- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).
CC		-I- SUBCELLULAR LOCATION: Nuclear (potential).
CC		-I- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA)
CC		-I- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL: S80870; AAB50691.1; -				
DR	InterPro: IPR003015; HLH_MYC				
DR	InterPro: IPR001092; HLH_dim				
DR	Pfam: PF00010; HLH; 1				
DR	SMART: SM00353; HLH; 1				
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; 1				
KW	Nuclear protein; Transcription regulation; Repressor.				
FT	DOMAIN 11	14	POLY-ALA.		
FT	DOMAIN 58	62	POLY-SER.		
FT	DOMAIN 83	99	POLY-GLY.		
FT	DOMAIN 174	179	POLY-GLY.		
FT	DOMAIN 204	217	POLY-GLY.		
FT	DNA_BIND 229	240	BASIC DOMAIN.		
FT	DOMAIN 241	282	HELIX-LOOP-HELIX MOTIF (
FT	DOMAIN 311	319	POLY-ALA.		
SQ	SEQUENCE 367 AA: 35905 MW: 6CAB3AF9F6E585F77				CRG64

Query Match 31.1%; Score 59; DB 1; Length 367;
Best Local Similarity 61.1%; Pred. No. 12;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```
QY      11 AARAGGGGGGGGIEPTL 28
          |   |||||:  | |
Db     88 AGGGGGGGGGGVSVPL 105
```

Query Match 31.38; Score 59.5; DB 1; Length 969;

RESULT 21

```

FXD2_HUMAN          STANDARD;          PRT;    497 AA.
ID  O60548;
AC  O60548;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  Forkhead box protein D2 (Forkhead-related protein FKHL17) (Forkhead-
DE  related transcription factor 9) (FREAC-9).
GN  FOXD2 OR FKHL17 OR FREAC9.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=98066765; PubMed=9403061;
RA  Ernström S., Betz R., Lagercrantz S., Larsson C., Eriksson S.,
RA  Cederberg A., Carlsson P., Enerbaeck S.;
RT  "Cloning and characterization of freac-9 (FKHL17), a novel kidney-
RT  expressed human forkhead gene that maps to chromosome 1p32-p34.";
RL  Genomics 46:78-85(1997).
RN  [2]
RP  REVISIONS.
RA  Enerbaeck S.;
RL  Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC  -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC  -!- SUBCELLULAR LOCATION: Nuclear.
CC  -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
CC  -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AF042832; AAC15421.1; -
CC  DR  HSP: O63245; 2HEF.
CC  DR  TRANSFAC; T02485; -.
CC  DR  MIM; 602211; -.
CC  DR  InterPro; IPR001766; Fork_head.
CC  DR  Pfam; PF00250; Fork_head.1.
CC  DR  PRINTS; PR00053; FORKHEAD.
CC  DR  SMART; SM00339; FH; 1.
CC  DR  PROSITE; PS00657; FORK_HEAD_1; 1.
CC  DR  PROSITE; PS00658; FORK_HEAD_2; 1.
CC  DR  PROSITE; PS00339; FORK_HEAD_3; 1.
CC  DR  DNA-binding; Nuclear protein; Transcription regulation.
CC  KW  DOMAIN 90 94 POLY-ALA.
FT  DOMAIN 101 104 POLY-ALA.
FT  DNA_BIND 126 217 FORK-HEAD.
FT  DOMAIN 247 250 POLY-ALA.
FT  DOMAIN 296 306 POLY-ALA.
FT  DOMAIN 398 409 POLY-GLY.
FT  DOMAIN 421 426 POLY-GLY.
FT  DOMAIN 442 445 POLY-ALA.
SQ  SEQUENCE 497 AA; 49007 MW; EFAF98DD216BE019 CRC64;

```

Query Match 31.1%; Score 59; DB 1; Length 497;
 Best Local Similarity 66.7%; Pred. No. 15;
 Matches 14; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

```

QY  4 PT--LRQCLAAARAGGGGGG 22
    |||||
DB  385 PTALLRQGLKTDAGGAGGGG 405

```

RESULT 22

Y967_TREPA

```

Y967_TREPA          STANDARD;          PRT;    517 AA.
ID  O83933;
AC  O83933;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Hypothetical protein TP0967.
GN  TP0967.
OS  Treponema pallidum.
OC  Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX  NCBI_TaxID=160;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=NICHOLS;
RX  MEDLINE=98332770; PubMed=9665876;
RA  Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA  Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA  Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA  Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA  McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA  Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA  Venter J.C.;
RT  "Complete genome sequence of Treponema pallidum, the syphilis
RT  spirochete.";
RL  Science 281:375-388(1998).
CC  -!- SIMILARITY: BELONGS TO THE TP096X FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AE001264; AAC65925.1; -
CC  DR  TIGR; TP0967; -.
CC  KW  Hypothetical protein; Complete proteome.
FT  DOMAIN 152 161 POLY-GLY.
SQ  SEQUENCE 517 AA; 56597 MW; E224976333989DF6 CRC64;

```

Query Match 31.1%; Score 59; DB 1; Length 517;
 Best Local Similarity 60.0%; Pred. No. 16;
 Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

```

QY  3 GPTLRQCLAAARAGGGGGG 22
    |.|.|.|.|.|.|.|.|
DB  141 GMTVTQPNCAAGGGGGG 160

```

RESULT 23

```

K1CJ_HUMAN          STANDARD;          PRT;    593 AA.
ID  K1CJ_HUMAN
AC  P13645;
DT  01-JAN-1990 (Rel. 13, Created)
DT  01-JUN-1994 (Rel. 29, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Keratin, type I cytoskeletal 10 (Cytokeratin 10) (CK 10).
GN  KRT10
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=89125611; PubMed=2464696;
RA  Rieger M., Franke W.W.;
RT  "Identification of an orthologous mammalian cytokeratin gene. High
RT  degree of intron sequence conservation during evolution of human
RT  cytokeratin 10.";
RL  J. Mol. Biol. 204:841-856(1988).
RN  [2]
RP  SEQUENCE OF 130-593 FROM N.A.
RX  MEDLINE=88122104; PubMed=2448602;

```

RA Darmon M.Y., Semat A., Darmon M.C., Vasseur M.;
RT "Sequence of a cDNA encoding human keratin No 10 selected according
RT to structural homologies of keratins and their tissue-specific
RT expression.";
RL Mol. Biol. Rep. 12:277-283(1987).
RN [3]
RX SEQUENCE OF 197-593 FROM N.A.
RX MEDLINE=92339897; PubMed=1378806;
RA Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P.,
RA Kisselev L.L.;
RT "Exons I and VII of the gene (Ker10) encoding human keratin 10
RT undergo structural rearrangements within repeats.";
RL Gene 116:245-251(1992).
RN [4]
RN SEQUENCE OF 180-184 AND 577-589.
RN TISSUE=Keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;
RX MEDLINE=93162043; PubMed=1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandekerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
RN [5]
RN VARIANT EHK HIS-156.
RX MEDLINE=92386600; PubMed=1381287;
RX Cheng J., Syder A.J., Yu Q.-C., Letai A., Faller A.S., Fuchs E.;
RA "The genetic basis of epidermolytic hyperkeratosis: a disorder of
RA differentiation-specific epidermal keratin genes.";
RL Cell 70:811-819(1992).
RN [6]
RN VARIANTS.
RX MEDLINE=92411228; PubMed=1371013;
RX Korge B.P., Gan S.-Q., McBridge O.W., Mischke D., Steinert P.M.;
RA "Extensive size polymorphism of the human keratin 10 chain resides in
RA the C-terminal V2 subdomain due to variable numbers and sizes of
RA glycine loops.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).
RN [7]
RN VARIANTS EHK HIS-156 AND SER-161.
RX MEDLINE=92376531; PubMed=1380725;
RX Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;
RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT hyperkeratosis.";
RL Science 257:1128-1130(1992).
RN [8]
RN VARIANTS EHK HIS-154; CVS-156; HIS-156; ASP-160 AND GLN-442.
RX MEDLINE=94136477; PubMed=7508181;
RX Chipev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L.,
RA Compton J.G., Bale S.J.;
RT "Preferential sites in keratin 10 that are mutated in epidermolytic
RT hyperkeratosis.";
RL Am. J. Hum. Genet. 54:179-190(1994).
RN [9]
RN VARIANTS EHK ARG-150; CVS-156 AND GLU-439, AND VARIANT SER-126.
RX MEDLINE=94216497; PubMed=7512983;
RX Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;
RA "Genetic mutations in the K1 and K10 genes of patients with
RA epidermolytic hyperkeratosis. Correlation between location and
RA disease severity.";
RL J. Clin. Invest. 93:1533-1542(1994).
RN [10]
RN VARIANT EHK ASN-160.
RX MEDLINE=94117868; PubMed=7507150;
RX Rothnagel J.A., Longley M.A., Holder R.A., Kuster W., Roop D.R.;
RA "Prenatal diagnosis of epidermolytic hyperkeratosis by direct gene
RA sequencing.";
RT J. Invest. Dermatol. 102:13-16(1994).
RN [11]
RN VARIANTS EHK PRO-156 AND SER-156.
RX MEDLINE=94117870; PubMed=7507152;
RX McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,
RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA Morley S.M.;
RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT congenital ichthyosiform erythroderma (BCIE).";
RL J. Invest. Dermatol. 102:24-30(1994).
RN [12]
RN VARIANT EHK THR-150.
RX MEDLINE=95059228; PubMed=7526210;
RX Paller A.S., Syder A.J., Chan Y.-M., Yu Q.-C., Hutton M.E., Tadini G.,
RA Fuchs E.;
RT "Genetic and clinical mosaicism in a type of epidermal nevus.";
RL New Engl. J. Med. 331:1408-1415(1994).
RN [13]
RN VARIANT AEI THR-446.
RX MEDLINE=99072665; PubMed=9856845;
RX Suga Y., Duncan K.O., Heald P.W., Roop D.R.;
RA "A novel helix termination mutation in keratin 10 in annular
RA epidermolytic ichthyosis, a variant of bullous congenital
RA ichthyosiform erythroderma.";
RL J. Invest. Dermatol. 111:1220-1223(1998).
RN [14]
RN VARIANT EHK SER-160.
RX MEDLINE=99215719; PubMed=10201536;
RX Arin M.J., Longley M.A., Anton-Lamprecht I., Kurze G., Huber M.,
RA Hohl D., Rothnagel J.A., Roop D.R.;
RT "A novel substitution in keratin 10 in epidermolytic hyperkeratosis.";
RL J. Invest. Dermatol. 112:506-508(1999).
CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.
CC -1- TISSUE SPECIFICITY: SEEN IN ALL SUPRABASAL CELL LAYERS INCLUDING
CC STRATUM CORNEUM.
CC -1- POLYMORPHISM: A NUMBER OF ALLELES ARE KNOWN THAT MAINLY DIFFER IN
CC THE GLY-RICH REGION (POSITIONS 490-560).
CC -1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF EPIDERMOLYTIC
CC HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC ICHTHYOSIFORM ERYTHRODERMA (BCIE)); A HEREDITARY SKIN DISORDER
CC CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC CORNEUM. AT BIRTH, AFFECTED INDIVIDUALS USUALLY PRESENT WITH
CC REDNESS, BLISTERS AND SUPERFICIAL EROSIONS DUE TO CYTOLYSIS.
CC WITHIN A FEW WEEKS, THE ERYTHRODERMA AND BLISTER FORMATION
CC DIMINISH AND HYPERKERATOSES DEVELOP. TRANSMISSION IS AUTOSOMAL
CC DOMINANT, BUT MOST CASES ARE SPORADIC.
CC -1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF ANNULAR EPIDERMOLYTIC
CC ICHTHYOSIS (AET), A DISTINCT PHENOTYPIC VARIANT OF EPIDERMOLYTIC
CC HYPERKERATOSIS. IT RESEMBLES CLINICAL AND HISTOLOGIC FEATURES OF
CC BOTH EPIDERMOLYTIC HYPERKERATOSIS AND ICHTHYOSIS BULLOSA OF
CC SIEMENS.
CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II
CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY IN
CC POSITIONS 513 TO 555.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14487; CAA32649.1; -
CC EMBL: M19156; AAA59468.1; -
CC EMBL: M77663; AAA59199.1; -
CC EMBL: L20218; AAB59438.1; -
CC EMBL: L20219; AAB59439.1; -
CC PIR: S02158; KRHU0.
CC Aarhus/Ghent-2DPAGE; 7405; IEF.
CC MIM: 148080; -
CC MIM: 113800; -
CC InterPro: IPR001664; IF.
CC InterPro: IPR002957; Keratin_I.
CC Pfam: PF00038; filament; 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U75355; AAB81130.1; -;
 DR HSP; P20393; IAO1;
 DR InterPro; IPR000536; Hormone_rec_lig.
 DR InterPro; IPR001723; Strdhormone_receptor.
 DR InterPro; IPR001628; zf-C4.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00105; zf-C4; 1.
 DR PRINTS; PR00398; STRDHORMONER.
 DR PRINTS; PR00047; STROIDFINGER.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00399; ZnF_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger.
 FT DOMAIN 1 300 MODULATING (POTENTIAL).
 FT DNA_BIND 301 366 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 301 321 C4-TYPE.
 FT ZN_FING 337 361 C4-TYPE.
 FT DOMAIN 454 674 HORMONE-BINDING (POTENTIAL).
 SQ SEQUENCE 757 AA; 83075 MW; C1511452ED3D7359 CRC64;

Query Match 31.1%; Score 59; DB 1; Length 757;
 Best Local Similarity 76.9%; Pred. NO. 21;
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 15 GGGGGGGIEGPT 27
 |||||
 DB 129 GGGGGGGVPGMT 141

RESULT 26

SSB_MYCLE
 ID SSB_MYCLE STANDARD; PRT; 168 AA.
 AC P46390; O53126;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Single-strand binding protein (SSB) (Helix-destabilizing protein).
 GN SSB OR ML2684 OR MLCB1913.20C.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97124199; PubMed=8969512;
 RA Fsihi H., de Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,
 RA Takiff H.E., Eiglmeyer K., Bergh S., Cole S.T.;
 RT "Gene arrangement and organization in a approximately 76 kb fragment
 RT encompassing the *oriC* region of the chromosome of *Mycobacterium*
 RT *leprae*."
 RL Microbiology 142:3147-3161(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR REPLICATION OF THE
 CC CHROMOSOME. IT IS ALSO INVOLVED IN DNA RECOMBINATION AND REPAIR
 CC (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SSB FAMILY.
 CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 137.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; L39923; AAB53120.1; ALT_FRAME.
 DR EMBL; AL022118; CAA17953.1; -;
 DR EMBL; AL583926; CAC32216.1; -;
 DR Leproma; ML2684; -;
 DR HSP; P02339; ILYG;
 DR InterPro; IPR000424; SSB.
 DR Pfam; PF00436; SSB; 1.
 DR PROSITE; PS00735; SSB_1; FALSE_NEG.
 DR PROSITE; PS00736; SSB_2; FALSE_NEG.
 KW DNA-binding; DNA repair; DNA replication; Complete proteome.
 FT DOMAIN 124 133
 SQ SEQUENCE 168 AA; 17700 MW; 077C62E430623658 CRC64;

Query Match 30.8%; Score 58.5; DB 1; Length 168;
 Best Local Similarity 52.0%; Pred. NO. 6.7;
 Matches 13; Conservative 3; Mismatches 4; Indels 5; Gaps 1;

QY 3 GPTLRQCL-----AARAGGGGGGG 22
 |||||
 DB 107 GPSLRYATAKVNKASRSGGGGGG 131

RESULT 27

HH3R_HUMAN
 ID HH3R_HUMAN STANDARD; PRT; 445 AA.
 AC Q9Y5N1; Q9H4K8; Q9GZX2;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Histamine H3 receptor (HH3R) (G protein-coupled receptor 97).
 GN HRH3 OR GPCR97.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Thalamus;
 RX MEDLINE=99278519; PubMed=10347254;
 RA Lovenberg T.W., Roland B.L., Wilson S.J., Jiang X., Pyati J.,
 RA Huvar A., Jackson M.R., Erlander M.G.;
 RT "Cloning and functional expression of the human histamine H3
 RT receptor."
 RL Mol. Pharmacol. 55:1101-1107(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
 RX MEDLINE=20568725; PubMed=11118334;
 RA Nakamura T., Itadani H., Hidaka Y., Ohta M., Tanaka K.;
 RT "Molecular cloning and characterization of a new human histamine
 RT receptor, *HRH4R*."
 RL Biochem. Biophys. Res. Commun. 279:615-620(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4; 5; 6 AND 7).
 RC TISSUE=Thalamus;
 RX MEDLINE=21181559; PubMed=11284713;
 RA Coye F., Guenin S.-P., Audinot V., Renouard-Try A., Beauverger P.,


```

RN  SEQUENCE FROM N.A.
RP  Birren B., Linton L., Nusbaum C., Lander E.;
RL  Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN  SEQUENCE OF 144-300 FROM N.A.
RX  MEDLINE=91257849; PubMed=1675198;
RA  D'Esposito M., Morelli F., Acampora D., Migliaccio E., Simeone A.,
RA  Boncinelli E.;
RT  EVX2, a human homeobox gene homologous to the even-skipped
RT  segmentation gene, is localized at the 5' end of HOX4 locus on
RT  chromosome 2.;
RL  Genomics 10:43-50(1991).
CC  -!- SUBCELLULAR LOCATION: Nuclear.
CC  -!- DEVELOPMENTAL STAGE: EXPRESSED DURING EARLY EMBRYOGENESIS AND
CC  NEUROGENESIS IN A BIPHASIC MANNER.
CC  -!- SIMILARITY: BELONGS TO THE EVEN-SKIPPED FAMILY OF HOMEBOX
CC  PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AC009336; -; NOT_ANNOTATED_CDS.
DR  EMBL; M59983; AAS52414.1; -
DR  EMBL; M59982; AAS52414.1; JOINED.
DR  HSSP; P14653; 1B72.
DR  MIM; 142991; -
DR  InterPro; IPR001356; HTH_repressr.
DR  InterPro; IPR001356; Homeobox.
DR  PRINTS; PR00024; HOMEBOX.
DR  PRINTS; PR00031; HTHREPRESSR.
DR  SMART; SM00389; HOX; 1.
DR  PROSITE; PS00027; HOMEBOX_1; 1.
DR  PROSITE; PS00071; HOMEBOX_2; 1.
KW  DNA-binding; Developmental protein; Homeobox; Nuclear protein.
FT  DNA_BIND 188 247 HOMEBOX.
FT  DOMAIN 294 301 POLY-ALA.
FT  DOMAIN 304 308 POLY-ALA.
FT  DOMAIN 346 351 POLY-ALA.
FT  DOMAIN 356 370 POLY-ALA.
FT  DOMAIN 373 378 POLY-ALA.
FT  DOMAIN 398 408 POLY-ALA.
FT  DOMAIN 413 434 POLY-GLY.
SQ  SEQUENCE 476 AA; 47799 MW; 6AA99041BA151C3F CRC64;

Query Match 30.5%; Score 58; DB 1; Length 476;
Best Local Similarity 76.9%; Pred. No. 18;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 10 LAARAGGGGGGG 22
Db 409 LGSRRGGGGGGG 421

RESULT 29
ONC2_HUMAN
ID ONC2_HUMAN STANDARD; PRT; 485 AA.
AC O95948;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE One cut domain family member 2 (ONECUT-2 transcription factor) (OC-2).
GN ONECUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

[1]
RN  SEQUENCE FROM N.A.
RP  MEDLINE=91115605; PubMed=9915796;
RA  Jacquemin P., Lannoy V., Rousseau G.G., Lemaigre F.P.;
RT  "OC-2, a novel mammalian member of the ONECUT class of homeodomain
RT  transcription factors whose function in liver partially overlaps with
RL  J. Biol. Chem. 274:2665-2671(1999).
CC  -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES THE TRANSCRIPTION
CC  OF A NUMBER OF LIVER GENES SUCH AS HNF3B.
CC  -!- SUBCELLULAR LOCATION: Nuclear.
CC  -!- SIMILARITY: CONTAINS 1 CUT DOMAIN.
CC  -!- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEBOX PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Y18198; CAB38253.1; -
DR  TRANSFAC; T03259; -
DR  MIM; 604894; -
DR  InterPro; IPR001350; CUT.
DR  InterPro; IPR001356; Homeobox.
DR  Pfam; PF02376; CUT; 1.
DR  Pfam; PF00046; homeobox; 1.
DR  SMART; SM00389; HOX; 1.
DR  PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.
DR  PROSITE; PS00071; HOMEBOX_2; 1.
KW  Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW  Activator.
FT  DNA_BIND 305 391 CUT.
FT  DNA_BIND 407 466 HOMEBOX.
FT  DOMAIN 18 37 POLY-GLY.
FT  DOMAIN 62 66 POLY-PRO.
FT  DOMAIN 75 82 POLY-ALA.
FT  DOMAIN 152 165 POLY-HIS.
FT  DOMAIN 298 303 POLY-SER.
SQ  SEQUENCE 485 AA; 52482 MW; AF21E052EFBE5DA1 CRC64;

Query Match 30.5%; Score 58; DB 1; Length 485;
Best Local Similarity 65.0%; Pred. No. 19;
Matches 13; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 15 GGGGGGGGGIEGPTLRQCLAA 34
Db 25 GGGGGGGGGGGPGHEQELLA 44

RESULT 30
BRNL_MOUSE
ID BRNL_MOUSE STANDARD; PRT; 495 AA.
AC P31361;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Brain-specific homeobox/POU domain protein 1 (BRN-1 protein).
GN POU3F3 OR OTF8 OR BRN1 OR BRN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92228768; PubMed=1565620;
RA Hara Y., Rovescalli C., Kim Y., Nirenberg M.;
RT "Structure and evolution of four POU domain genes expressed in mouse
RT brain.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3280-3284(1992).
CC -!- SUBCELLULAR LOCATION: Nuclear.

```


Wed Oct. 9 10:29:52 2002

```

CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
CC TO CLASS-3 POU.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88299; AAA39960.1; -
CC PIR; S31223; S31223.
CC HSP; P14859; IOCT.
CC MGD; MGI:102564; Pou3f3.
CC InterPro; IPR001356; Homeobox.
CC InterPro; IPR000327; POU.
CC Pfam; PF00046; homeobox; 1.
CC Pfam; PF00157; pou; 1.
CC PRINTS; PR00028; POUDOMAIN.
CC ProDom; PD000583; POU; 1.
CC SMART; SM00389; Hox; 1.
CC SMART; SM00352; POU; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC PROSITE; PS00035; POU_1; 1.
CC PROSITE; PS00465; POU_2; 1.
CC PROSITE; PS00465; POU_2; 1.
CC KW Nuclear protein; DNA-binding; Homeobox.
CC FT DOMAIN 28 49 POLY-GLY.
CC FT DOMAIN 101 112 POLY-ALA.
CC FT DOMAIN 186 201 POLY-ALA.
CC FT DOMAIN 267 291 HIS-RICH.
CC FT DOMAIN 313 383 POU.
CC FT DNA_BIND 401 460 HOMEBOX.
CC SQ SEQUENCE 495 AA; 50012 MW; 77B02E890C9A014 CRC64;

Query Match 30.5%; Score 58; DB 1; Length 495;
Best Local Similarity 91.7%; Pred. No. 19;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 AARAGGGGGGG 22
Db 26 AAGAGGGGGGG 37

```

Search completed: October 9, 2002, 09:00:14
 Job time : 5.3831 secs

